5-2010

A NEW TUMOR SUPPRESSOR GENE CANDIDATE REGULATED BY THE NON-CODING RNA PCA3 IN HUMAN PROSTATE CANCER

Alessandro K. Lee

A NEW TUMOR SUPPRESSOR GENE CANDIDATE REGULATED BY THE NON-CODING RNA PCA3 IN HUMAN PROSTATE CANCER

by

Alessandro Kelien Lee

APPROVED:

Wadih Arap, M.D., Ph.D., Supervisory Professor

Renata Pasqualini, Ph.D.

Neal Pellis, Ph.D.

Frank Marini III, Ph.D.

Oliver Bogler, Ph.D.

APPROVED:

Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

A NEW TUMOR SUPPRESSOR GENE CANDIDATE REGULATED BY THE NON-CODING RNA PCA3 IN HUMAN PROSTATE CANCER

by

Alessandro Kelien Lee

APPROVED:

_____
Wadih Arap, M.D., Ph.D., Supervisory Professor

_____
Renata Pasqualini, Ph.D.

_____
Neal Pellis, Ph.D.

_____
Frank Marini III, Ph.D.

_____
Oliver Bogler, Ph.D.

APPROVED:

_____
Dean, The University of Texas
Graduate School of Biomedical Sciences at Houston

A NEW TUMOR SUPPRESSOR GENE CANDIDATE REGULATED BY THE NON-CODING RNA PCA3 IN PROSTATE CANCER

A DISSERTATION PRESENTED TO THE FACULTY OF
THE UNIVERSITY OF TEXAS HEALTH SCIENCE CENTER AT HOUSTON
AND THE UNIVERSITY OF TEXAS M. D. ANDERSON CANCER CENTER
GRADUATE SCHOOL OF BIOMEDICAL SCIENCES
IN PARTIAL FULFILLMENT OF THE REQUIREMENTS
FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY

BY

ALESSANDRO KELIEN LEE

Houston, Texas
May, 2010

ACKNOWLEDGMENT

The thesis work has been carried out at the University of Texas M. D. Anderson Cancer Center in Houston, Texas. I would like to thank my supervisor Dr. Wadih Arap and my advisor Dr. Renata Pasqualini for the opportunity, guidance, and privilege of working with them, and for all the invaluable help that they provided throughout my graduate studies.

I would also like to acknowledge the time and effort of my supervisory committee members, Drs. Neal Pellis, Frank Marini III, and Oliver Bogler for their essential advises, and Drs. Ricardo Brentani, Helen Sage, and Nora Navone for kindly agreeing to read this work.

A special acknowledgment is reserved for Drs. Emmanuel Dias-Neto and Ahmad Salameh, for all their help and mentorship, and without whom this work could not be realized.

This thesis is dedicated to my loving and supportive family, my wife Viviana for all her love, patience and encouragement, my parents Joseph and Emanuela, and to my son Gael, who inspired and motivated the realization of this work.

ABSTRACT

A NEW TUMOR SUPPRESSOR GENE CANDIDATE REGULATED BY THE NON-CODING RNA PCA3 IN HUMAN PROSTATE CANCER

Publication No._____

Alessandro Kelien Lee

Supervisory Professor: Wadih Arap, M.D., Ph.D.

Prostate cancer is the second leading cause of cancer-related death and the most common non-skin cancer in men in the USA. Considerable advancements in the practice of medicine have allowed a significant improvement in the diagnosis and treatment of this disease and, in recent years, both incidence and mortality rates have been slightly declining. However, it is still estimated that 1 man in 6 will be diagnosed with prostate cancer during his lifetime, and 1 man in 35 will die of the disease.

In order to identify novel strategies and effective therapeutic approaches in the fight against prostate cancer, it is imperative to improve our understanding of its complex biology since many aspects of prostate

cancer initiation and progression still remain elusive. The study of tumor biomarkers, due to their specific altered expression in tumor versus normal tissue, is a valid tool for elucidating key aspects of cancer biology, and may provide important insights into the molecular mechanisms underlining the tumorigenesis process of prostate cancer.

PCA3, is considered the most specific prostate cancer biomarker, however its biological role, until now, remained unknown. PCA3 is a long non-coding RNA (ncRNA) expressed from chromosome 9q21 and its study led us to the discovery of a novel human gene, PC-TSGC, transcribed from the opposite strand and in an antisense orientation to PCA3.

With the work presented in this thesis, we demonstrate that PCA3 exerts a negative regulatory role over PC-TSGC, and we propose PC-TSGC to be a new tumor suppressor gene that contrasts the transformation of prostate cells by inhibiting Rho-GTPases signaling pathways. Our findings provide a biological role for PCA3 in prostate cancer and suggest a new mechanism of tumor suppressor gene inactivation mediated by non-coding RNA. Also, the characterization of PCA3 and PC-TSGC led us to propose a new molecular pathway involving both genes in the transformation process of the prostate, thus providing a new piece of the jigsaw puzzle representing the complex biology of prostate cancer.

TABLE OF CONTENTS

LIST OF ILLUSTRATIONS

LIST OF TABLES

CHAPTER 1:

INTRODUCTION

1.1 Functions and biological significance of non-coding RNAs in eukaryotes: an overview

In the past twenty years, a significant number of discoveries in the RNA molecular biology field led the scientific community to the revision of the central dogma of biology: “DNA makes RNA makes protein”. This classical view is no longer entirely true as it is now clear that RNA is not “just” a molecule that translates the genetic information into proteins (Carnici *et al.,* 2008). Up to recent years, apart from mRNAs the only other RNA classes that were recognized to have a biological function were rRNA, tRNA and the RNA components of the spliceosome (snRNA), however the scientific community was still unaware of the enormous role of RNA in several other aspects of cellular life.

After the sequencing of the human genome, scientists realized that only a small percentage of the whole genome is actually transcribed into coding sequences that eventually are translated into proteins, and the rest of the DNA that does not contain coding regions was merely referred to as “junk DNA”. Also, transcription of non-coding regions was considered “transcriptional noise” (Louro *et al.,* 2009). No other definition could have been less accurate than these two.

After the discovery of the first miRNA, *lin-4* (Lee *et al.,* 1993), scientists realized that RNA could have important regulatory functions also in eukaryotes, considering that antisense RNA was already known to

regulate gene expression in bacteria (Brantl, 2007). However, only in the last 10 years, after the discovery of RNA interference, biologists really learned that the majority of the genome is actually functionally transcribed, and that non-coding RNAs are important in many aspects of cellular life, with particular attention to gene expression regulation (Amaral *et al.,* 2008).

Ever since, many different classes of regulatory RNA have been discovered and most likely many more are still unknown (Storz *et al.,* 2005). A precise classification of all the RNA classes is not yet available. Thus scientists generally classify them according to their size, biogenesis, interacting partner proteins and function (Amaral *et al.,* 2008). However the boundaries for each class is not always well defined and RNA nomenclature is not always accurate in relation to all the several classes that have been discovered so far.

In general, RNA classes involved in the regulation of gene expression can be divided into two main groups:

- Small non coding RNAs
- Long non coding RNAs

The distinction between small and long ncRNAs is just based on the fact that lncRNAs are arbitrarily considered to be longer than 200 nt,

on the basis of a practical cut-off in RNA purification protocols that excludes small RNAs (Storz *et al.,* 2005).

1.2 Small non-coding RNAs

Small ncRNAs are the most studied and well-characterized classes of regulatory RNAs. Among them the most important are:

Micro RNAs (miRNA): miRNAs are short RNAs approximately 22 nt long that usually act as regulators of translation but can also be involved in RNA degradation. miRNA may be expressed at high levels, up to tens of thousands of copies per cell, and play important regulatory roles by controlling hundreds of mRNA targets. They are transcribed by RNA polymerase II and transcripts are capped and polyadenylated. Half of all miRNAs are identified in clusters that are transcribed as polycistronic primary transcripts. Processing of the primary transcripts depends on the miRNA sequence folding into a stem-loop structure. A typical primary miRNA is cleaved by the complex Drosha-DGCR8, which is a RNase III family endonuclease, containing two RNase III domains and a dsRNA binding domain. The cleavage occurs 11nt from the base of the stem structure and the resulting pre-miRNA has a 5'-phosphate and a 3' 2nt overhang. An alternative pathway uses splicing of pri-miRNA transcripts, in this case also called mirtrons, to liberate introns that precisely mimic the structural features of pre-miRNA. After this first processing step, pre-miRNA are exported from the nucleus with a GTP-dependent transport mechanism to the cytoplasm, where they undergo a second processing

step operated by Dicer, which is another RNase III enzyme that excises the terminal loop from the pre-miRNA to create a mature miRNA of approximately 22nt. The miRNA processing machinery produces final miRNA duplexes molecules with very exact ends, and both Drosha and Dicer exploit RNA binding domains or accessory protein (DGR8) that allow the two enzymes to determine the cleavage sites in a ruler-like fashion. Next, the miRNA RISC (RNA Induced Silencing Complex) is formed by loading of one strand of the miRNA duplex onto a protein of the Argonaute family (Ago). Ago proteins are crucial effectors for the final silencing function of a miRNA and they can discriminate the miRNA strand to be loaded based on the relative thermodynamic stability of the miRNA ends: the strand with the lowest stability at the 5' end will be loaded into the RISC, while the other strand (passenger strand) is discarded, however this rule is not absolute and many aspects of miRNA unwinding and loading into RISC are unclear.

Once assembled, the miRNA acts as an adaptor for the RISC to specifically recognize and regulate target mRNAs. Generally, miRNA binding sites in animal mRNAs lie in the 3'UTR and usually are present in multiple copies. The recognition of the binding sites involves the nucleotides 2-8 of the miRNA, which represent the "seed" region. The degree of miRNA-mRNA complementarity determines the kind of regulatory mechanism adopted by the RISC: Perfect complementarity allows the cleavage of the mRNA strand by Ago. This cleavage is

operated by a domain of the Ago protein known as PIWI domain that acts as an RNase H endonuclease. Once the mRNA is cleaved, the exposed, unprotected ends are then susceptible to the attack of exonucleases. On the other hand, central mismatches of the miRNA-mRNA pair exclude cleavage, instead they promote repression of mRNA translation or removal of poly-A tails. In these cases, different mechanisms have been proposed:

- The RISC may promote the premature ribosome dissociation from mRNA.
- The RISC may compete with the elongation factor eIF4E for binding to the mRNA capping structure, therefore inhibiting translation initiation.
- The RISC may block association of the ribosomal 60S subunit with the 40S preinitiation complex.
- The RISC may stimulate the de-adenylation of the mRNA tail, leading to mRNA degradation.

It is possible that all these mechanisms may co-exist and even be synergistic. It is also very important to note that miRNA are able to stimulate protein translation under certain conditions. It seems that some miRNA inhibit translation in proliferating cells while they can stimulate translation during cell starvation. In another case, the binding of the miRNA at the 5' or 3'UTR regions determines the stimulation or repression

of translation, respectively. However, these do not represent general rules and also the mechanisms for stimulation of translation by miRNAs are not yet understood (Carthew *et al.,* 2009; Wu *et al.,* 2008).

Small interfering RNAs (siRNA): siRNA processing and silencing mechanisms resemble those of miRNA in many steps of the RNA interference process. The main differences between siRNA and miRNA lie in the biogenesis and in the target mRNAs of these two molecules.

In general, the term “siRNA” should include several sub-classes of small interfering RNAs, however “siRNA” is usually associated with a process that involves the cleavage of target mRNA molecules by a RISC complex loaded with a perfectly matching siRNA guide strand. The typical siRNA pathway is initiated in the cytoplasm when exogenous dsRNA enters the cell. Similarly to miRNAs, long dsRNA are processed by Dicer, which creates small dsRNAs with a 5’-phosphate 2nt overhang at the 3’ ends, then the 21-23nt strand with lower relative thermodynamic stability at the 5’ end will be loaded onto an Ago protein as the siRNA guide strand. At this point the RISC effector complex will be formed. With perfect base pairing and formation of an A-form helix structure between the siRNA guide strand and its target mRNA, the RISC cleaves (with the PIWI domain of Ago) its target 10-11 nt from the 5’ end of the guide siRNA strand, and the complex is then recycled for the next round of target

mRNA cleavage. mRNAs cleaved by the RISC subsequently are degraded by cellular exonucleases.

In some organisms, a siRNA amplification pathway was detected: a RNA-dependent RNA polymerase (RdRP) synthesized dsRNA from RNA cleaved by the RISC, and the synthesized dsRNA are used as siRNA precursors (Carthew *et al.,* 2009; Wu *et al.,* 2008).

Since endogenous siRNA pathways initially were not detected, originally this mechanism was thought to represent a mean of defense against exongenous dsRNA introduced into a cell, such as a virus. However, more recently siRNA regulatory mechanisms have been identified also for endogenous sequences in different organisms, leading to the discovery of new sub-classification and new mechanisms of action of siRNAs (Farazi *et al.,* 2008):

Piwi-interacting RNAs (piRNA): piRNAs are a class of small RNA (25-29nt) involved in RNA silencing or transposable elements in the germline. piRNA biogenesis is Dicer-independent and they are generally processed from single-stranded primary transcripts that are clustered into a limited number of genomic loci, where only one strand encodes piRNAs mainly.

Piwi-related proteins compose a sub-family belonging to the Ago family of effector proteins and they are specifically expressed in the

germline. Piwi proteins associate with piRNAs that act as guides in silencing target RNA. It is hypothesized that piRNAs are involved in the silencing of transposable elements during spermatogenesis (Kawaji *et al.,* 2008; Chu *et al.,* 2007).

Repeat-associated RNAs (rasiRNA): rasiRNAs originate from repetitive sequences such as transposons and are found in both the sense and antisense strands of transposable elements. rasiRNAs are suggested to repress the transposable elements themselves or mRNAs bearing sequences complementary to them. Also, rasiRNAs are involved in transcriptional silencing through chromatin modifications. Repetitive sequences frequently are found in chromatin domains, and it is suggested that rasiRNAs contribute to the regulation of chromatin status (Kawaji *et al.,* 2008; Chu *et al.,* 2007).

Heterochromatic siRNAs (hcRNAs): hcRNAs are a particular class of siRNAs that derive from long dsRNA precursors that are transcribed from genomic repeat regions such as centromeric repeats. They are involved in the regulation of heterochromatin structures and mediate transcriptional gene regulation. They are loaded onto a particular Ago-containing effector complex known as RITS (RNA Induced Transcriptional Silencing) and they guide the RITS to centromeric repeats by hcRNA recognition of nascent transcripts. The RITS is then involved in the

nucleation and maintenance of heterochromatin by recruiting histone methyltransferases. Engagement of nascent transcripts by the RITS also activates the RNA dependent RNA polymerase complex (RDRC) that uses its RdRP subunit to generate secondary hcRNAs. Also, in plants in addition of histone methylations, the RITS complex directly recruits DNA methyltransferases that contribute to the heterochromatin state (Kawaji *et al.,* 2008; Chu *et al.,* 2007).

Trans-acting siRNAs (tasiRNA): tasiRNA is a class of siRNA that targets other genes rather than the gene producing the siRNA itself. It is found only in plants and it is derived from an mRNA cleaved by a miRNA. A cleaved product of the mRNA is then used as a template to synthesize dsRNA by a RdRP, and the resulting product then triggers RNA silencing to repress other genes (Kawaji *et al.,* 2008; Chu *et al.,* 2007).

1.3 Long non-coding RNAs

Long non coding RNAs, the second main group of gene expression regulatory RNAs, were first described during large-scale sequencing of full-length mouse cDNA libraries (Ponting *et al.,* 2009). Ever since, it has become apparent that a significant portion of the transcriptome has little or no protein-coding capacity and it is now known that the extent of non-coding sequence transcription is at least four times greater than coding sequence (Louro *et al.,* 2009; Mendes Soares *et al.,* 2006; Amaral *et al.,* 2005). Long ncRNA often overlaps with, or intersperses between, multiple coding and non-coding transcripts and they are observed in a wide range of different tissues. These observations raise the possibility that genomic sequences can be transcribed into a range of sense and antisense, coding and non-coding transcripts. Most long ncRNAs tend to be transcribed away from the 5' or 3' ends of genes. Nevertheless, transcription of long ncRNAs has been found to be more concentrated near the promoters and initial exons and introns of genes (Mercer *et al.,* 2009). So far, a classification of long ncRNA is only based on their genomic proximity to protein coding genes, including overlapping, antisense, bidirectional or intronic long ncRNAs, even though many transcripts resist classification into any particular category, and instead exhibit a combination of these qualities.

Alongside tissue specificity, long ncRNAs transcription seems to be spatially restricted by subcellular expression preferences. For example, intronic long ncRNA expression appears predominantly nuclear, even though exceptions exist. Also, some long ncRNAs show unusual or unique localization patterns, thereby classifying entirely new subcellular compartments (Ponting *et al.,* 2009; Mercer *et al.,* 2009).

Long ncRNAs genereally show low sequence conservation, however they have conserved secondary structures and splicing patterns suggesting different roles of long ncRNA in gene expression regulation. Transcription of long ncRNAs may regulate the expression of genes in close genomic proximity (cis-acting regulation) and to target distant transcriptional activators or repressors by a variety of mechanisms involving chromatin remodeling, transcriptional and post-transcriptional regulation (Mercer *et al.,* 2009; Yazgan *et al.,* 2007).

Long ncRNAs may mediate epigenetic changes by recruiting chromatin remodeling complexes to specific genomic loci, where they define chromatin domains of differential histone methylation and RNA polymerase accessibility. These activities may be very important during development and for epigenetic phenomena such as imprinting and X chromosome inactivation. Another mechanism of chromatin remodeling involves the inhibition of the histone acetyltransferase activities of CREB binding proteins and p300 in order to silence target genes expression.

Apart from determining DNA accessibility by recruiting of chromatin remodeling complexes, it is suggested that the process of transcription of a long ncRNA per se, rather than the product of transcription, actually regulates transcription of target genes. These regulations are either positive or negative. For example, the active transcription of a long ncRNA may disrupt chromatin structure causing progressive opening of chromatin and promote the accessibility of protein-coding genes to RNA polymerase. Conversely, a negative regulation may result from collisions between transcriptional machineries that are processing along adjacent sequences. When elongation of one transcriptional event proceeds through a promoter sequence it suppresses this sequence's ability to initiate a second transcriptional event (Ponting *et al.,* 2009; Munroe *et al.,* 2006).

Long ncRNAs also act as co-factors to modulate transcription factor activity. For example, they recruit the binding and action of transcription factors to enhancers and then induce expression of adjacent protein-coding genes. Also, long ncRNAs may regulate RNA polymerase II activity through direct binding to DNA sequences. For example, transcribed long ncRNA forms a triplex with adjacent promoter regions and therefore inhibits binding of transcription factors and the formation of the pre-initiation complex required for gene transcription (Yazgan *et al.,* 2007)..

Long ncRNA also effects global changes by interacting with basal components of the RNA polymerase II transcription machinery. Such long

ncRNAs are typically transcribed by RNA polymerase III, thereby decoupling their expression from the RNA polymerase II–dependent transcription reaction they regulate. Other long ncRNAs regulate transcription indirectly by controlling the subcellular localization of transcription factors to the cytoplasm and preventing them to access the nucleus, maybe by interactions with nuclear transport factors (Yazgan *et al.,* 2007).

Many long ncRNA are transcribed antisense to corresponding genes and the ability of long ncRNA to recognize complementary sequences also allows highly specific interactions that regulate different steps in the post-transcriptional processing of mRNAs, including their splicing, editing, transport, translation and degradation. Antisense long ncRNAs can mask key cis elements in mRNA by the formation of RNA duplexes. This leads to alternative splicing event of the target mRNA by masking or exposing differential splice sites, or to the RNA editing of the annealed RNA sequences by specific double-strand RNA binding proteins such as ADAR1 and ADAR2. Another possibility of control by long ncRNA is that antisense long ncRNA stabilizes protein-coding RNA transcribed from the same genomic loci, which otherwise would undergo degradation (Munroe *et al.,* 2006; Katayama *et al.,* 2005).

Finally, the annealing of ncRNA may possibly target protein effector complexes to the sense mRNA transcript in a manner analogous to the targeting of the RISC to mRNA by miRNAs and siRNAs. RNA duplexes resulting from the annealing of complementary transcripts or even of long ncRNAs with extended internal hairpins may be processed into endogenous siRNA to silence gene expression, raising the possibility that long ncRNA feed into RNA interference silencing pathways. Some long ncRNAs transcripts are also spliced, polyadenylated and exported into the cytoplasm; their functions is not fully understood but it is also interesting to ask how an RNA that is spliced and exported from the nucleus evades destruction by the nonsense mediated decay pathway (Ponting *et al.,* 2009; Mercer *et al.,* 2009).

There are probably many other functions of long ncRNAs that have not been proposed yet and this particular field of research is now in rapid expansion.

1.4 Long ncRNAs in prostate cancer

A class of long ncRNAs that maps to genomic regions corresponding to intronic sequences located on their opposite DNA strand (intronic-antisense ncRNAs), correlates with a number of events, including tumor grade in prostate cancer (Louro *et al.*, 2007; Nakaya *et al.*, 2007; Reis *et al.,* 2004). Several lines of evidence indicate that intronic-antisense ncRNA are also involved in the regulation of expression of the corresponding sense gene (Faghihi *et al.,* 2008; Scheele *et al.,* 2007; Louro *et al.*, 2007; Katayama *et al.,* 2005; Munroe *et al.,* 2006; Tasheva *et al.,* 1995). Also intronic-antisense ncRNA respond to androgen stimulation (Louro *et al.,* 2007). Androgens are implicated in the regulation of several physiological and pathological processes in both normal and neoplastic prostate cells.

These evidence suggests that intronic-antisense ncRNA may play a key role in the tumorigenesis process of PCa, possibly by regulating the gene expression levels as well as splicing variants of their corresponding sense gene.

In particular, PCA3 is a ncRNA that is specifically enhanced in human prostate cancer and appears to be a promising biomarker in this tumor. The expression of PCA3 correlates with a cancerous state of the prostate and is prostate tissue-specific. Furthermore, PCA3 is responsive to androgen stimulation and its cellular location is mainly confined to the

nucleus (Schalken *et al.,* 2003; de Kok *et al.,* 2002). However, though PCA3 can be easily considered as the most important ncRNA ever identified in PCa, and its significance as a new PCa biomarker bears enormous potential for the improvement of current methodologies in PCa diagnosis, functional characterization of PCA3 and potential biological roles still remain largely undetermined.

1.5 Pitfalls in PCa diagnosis: current challenges and new approaches

One of the most important challenges in the management and treatment of prostate cancer is efficient prediction of disease outcome. Current diagnostic methodologies do not allow a precise risk assessment for the development and recurrence of aggressive prostate cancer (Troyer *et al.,* 2004; Reynolds *et al.,* 2007). Digital Rectal Examination (DRE) and monitoring of Prostate Specific Antigen (PSA) serum levels are the current standard assays that are employed for an initial assessment of prostate conditions, while a definitive diagnosis of prostate cancer is only determined upon the evaluation of prostate biopsies. Specifically, the severity of the disease is determined upon histological analysis with the assignment of a combined score according to the Gleason grading system (Gleason *et al.,* 1974). The Gleason score is a combination of two grades on a scale from 1 to 5 that represent the most common and second most common patterns of cellular differentiation in a given specimen.

Although the introduction of PSA screenings has significantly improved the ability of detecting prostate malignancies, and the Gleason's scoring system has provided for a systematic classifications of tumors based on their histological features, the current diagnostic procedures are far from optimal, and this is mainly due to two significant limitations. First, PSA serum level has been proven to lack specificity for a cancerous state

versus other pathological conditions of the prostate such as prostatitis and Benign Prostatic Hyperplasia (BPH) (Troyer *et al.,* 2004; Reynolds *et al.,* 2007). It has been shown that only 25% to 35% of men with elevated PSA serum levels actually develop tumors (Smith *et al.,* 1997). This high rate of false positives results in overdiagnosis and exposure of the patients to unnecessary risks and complications associated to invasive procedures (biopsies) and collateral effects of prostate cancer treatment (hormonal, radiation and surgical therapies). Second, a low Gleason score (2 to 6) is assigned to well differentiated cell populations and usually is indicative of a favorable prognosis, while a high Gleason score (8 to 10) is assigned to poorly differentiated cell populations and usually is indicative of an aggressive and recurrent disease. However, research on diagnostic methodologies for prostate cancer shows that up to 45% of biopsies in a given sample population are classified as Gleason 7 (Sim *et al.*, 2007; Nguyen *et al.,* 2009), for which the prognosis is extremely unpredictable. This situation translates into an inability to discriminate between patients that need intensive treatment or even to be enrolled in experimental clinical trials, from patients who do not need any therapy and would actually be harmed by the side effects related to any treatment.

In this scenario, there was a great deal of effort in the past years to improve the tools that clinicians currently use for the diagnosis and risk assessment of prostate cancer, and particular attention has been given to the discovery of better diagnostic and prognostic biomarkers. By

extensively exploiting genomics, proteomics and bioinformatics techniques, the scientific community has identified and validated a set of biomarkers that bear a great potential for the improvement of current diagnostic procedures (Sardana *et al.,* 2008; Makarov *et al.,* 2008; Parekh *et al.,* 2007). One of these markers is PCA3, a long non-coding RNA (ncRNA) transcribed from chromosome 9q21 (Bussemaker *et al.,* 1999).

Over the years, PCA3 has been extensively validated as a gene specifically up-regulated in prostate cancer and its diagnostic potential as a cancer biomarker has been established (Schalken *et al.,* 2003; de Kok *et al.,* 2002). The literature is constantly updated with reports on the applicability of PCA3 testing to a clinical setting and several diagnostic assays that rely on the detection of PCA3 in biological fluids have been developed (van Gills *et al.,* 2007; de la Taille *et al.,* 2007; Groskopf *et al.,* 2006). Approximately ten years of research on PCA3 led to the establishment of this gene as a bona fide prostate cancer biomarker, however, a function for such an important gene and its role in the biology of prostate cancer is yet to be uncovered.

1.6 Aims and Objectives

To gain insight into the function of PCA3 and to identify corresponding antisense transcripts as yet unrecognized, we conducted a comprehensive analysis of partial cDNA sequences that flank *PCA3* on its opposite DNA strand. At this genomic locus, we found a transcript that spans ~300kb and identified a human gene antisense to *PCA3*: *PCA3 Controlled - Tumor Suppressor Gene Candidate* (*PC-TSGC*). We cloned and expressed two distinct splicing isoforms of *PC-TSGC*, and showed that expression of PC-TSGC is regulated by its intronic-antisense ncRNA PCA3 in human prostate cancer cells, data providing a potential function for this important biomarker. We prove that *PC-TSGC* encodes for a new protein that interacts with two partners: RhoA and nm23-H1, and demonstrate that decreased levels of PC-TSGC result in increased activation of AKT and ERK1/2 signaling pathways. Finally, our data show that PC-TSGC prevents anchorage-independent cell growth of prostate cancer cells, as well as tumor growth *in vivo*.

Taken together, we propose for the first time a role for the PC-TSGC and PCA3 interplay in the biology of human prostate cancer, and we show that these newly discovered PCA3- and PC-TSGC- mediated mechanisms yield insights and possible targets for therapeutic intervention in prostate cancer.

CHAPTER 2:

IDENTIFICATION OF PC-TSGC

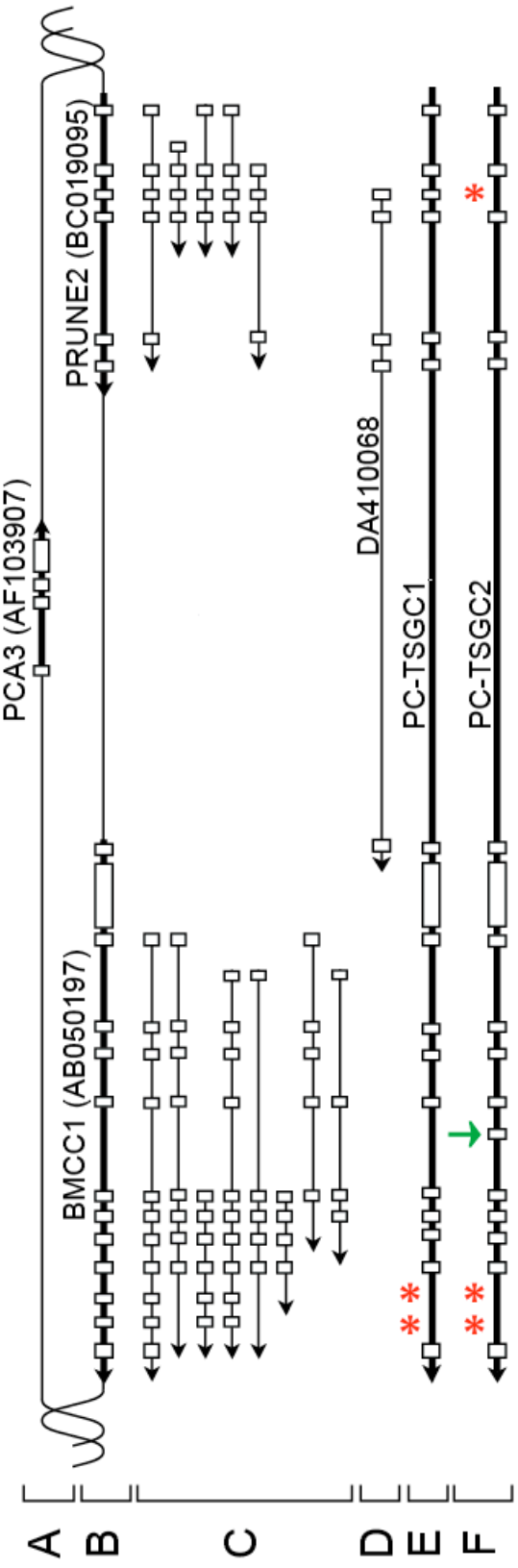
2.1 Analysis of chromosome 9q21 genomic region

The ncRNA PCA3 (GenBank accession number: AF103907) is transcribed from chromosome 9q21 into several isoforms. The largest is 4kb (Bussemakers *et al.,* 1999) (Figure 2.1A). Two distinct mRNAs on both sides of PCA3 are transcribed from the opposite DNA strand, in an antisense orientation to PCA3 (Figure 2.1B). The mRNA located downstream to PCA3 is transcribed from a gene named PRUNE2 (GenBank accession number: BC019095) (Strausberg *et al.,* 2002), while the mRNA upstream to PCA3 has actually been assembled from several cDNA fragments and referred to as BMCC1 (BCH-motif-containing molecule at the C-terminal region 1, GenBank accession number: AB050197) (Machida *et al.,* 2006). PRUNE2 and BMCC1 are annotated as full-length genes, and several mRNAs as well as spliced Expressed Sequenced Tags (ESTs) from the GenBank clearly align in correspondence to both sequences, marking the boundaries of their transcriptional start and end sites (Figure 2.1C). In this location, we noticed one single EST (GenBank accession number: DA410068) (Kimura *et al.,* 2006) that spans almost 140,000 base pairs, and specifically a genomic region including more than half of the PRUNE2 sequence, the full PCA3 sequence and the 5' end of BMCC1 (Figure 2.1D), alternatively indicating that the BC019095 and AB050197 sequences could actually be part of a single transcriptional unit. In this case, PCA3 would be then

transcribed from a genomic region that could be identified as an intron of a new gene not yet annotated. Recent studies have shown that among the overall expression of long ncRNAs in prostate cancer, a significant portion of these genes are transcribed from intronic regions in an antisense orientation to the corresponding sense gene (Louro *et al.,* 2007). These observations reinforced our reasoning that the annotated sequences identifying PRUNE2 and BMCC1 are incomplete sequences that are instead part of a single transcriptional unit spanning ~300kb. In this case, such transcript would extend from the 5' UTR region of PRUNE2 to the 3' UTR region of BMCC1, covering a significant intronic region that includes PCA3 and therefore classifying PCA3 as a intronic-antisense ncRNA.

Figure 2.1

Schematic representation of chromosome region 9q21.13 (not to scale). (*A* and *B*) Transcribed RNA sequences (black arrows) and exons (white boxes) are represented. Arrow heads indicate the orientation of the transcripts. (*C*) Representative ESTs correspond to the BC019095 and to the AB050197 sequences. (*D*) The EST DA410068 links the two EST clusters flanking PCA3. (*E* and *F*) Full-length PC-TSGC isoforms are shown: the alternative splice sites, are highlighted by red stars (missing exons) and by the green arrow (new exon).


PCA3 (AF103907)
BMCC1 (AB050197)
PRUNE2 (BC019095)
DA410068
PC-TSGC1
PC-TSGC2
A
B
C
D
E
F

2.2 Isolation and Cloning of PC-TSGC

To verify the expression of this single transcript we designed several forward primers targeted along the BC019095 sequence (PRUNE2) and reverse primers targeted along the AB050197 sequence (BMCC1). We tested different combinations of forward and reverse primers on cDNA samples prepared from a variety of cell lines and tissue samples. In brief, we outlined RT-PCR experiments to amplify several regions of the putative full-length gene, resulting in size of amplicons ranging from 0.1kb up to 12kb, where a 12kb PCR product would match the full-length transcript. To accomplish this, forward primers annealing to exon 1, exon 2, and exon 5 of PRUNE2 were used in combination with reverse primers annealing to exon 1, exon 2 and exon 13 of BMCC1. We screened by RT-PCR a panel of RNA samples from commercially available prostate cancer cell lines (LNCaP, PC3, DU145), lung cancer cell line (A549K), and glioma cell line (U87) considering that almost all of the ESTs aligning to either PRUNE2 or BMCC1 are derived from lung or brain tissues. In addition, we tested RNA of cell lines derived from other tissues such as kidney (HEK293) and soft tissue sarcoma (KS1767). Then, we included commercially available human RNA from prostate tumor as well as from other normal tissues (liver, kidney, breast, lung, pancreas, spleen and testes). Finally, we screened RNA extracted from human prostate tumor tissue samples and prostate tumor xenografts

models, including the MDA-PCa-118a, MDA-PCa-118b, and MDA-PCa-133, which are prostate cancer tumor graft established at the David H Koch Center, The University of Texas M. D. Anderson Cancer Center.

From this screening, we were able to obtain several RT-PCR products of the expected size from most of the screened cDNA samples, however the amplification of the full-length 12 kb cDNA fragment was accomplished only from a cDNA sample derived from the prostate cancer tumor graft MDA-PCa-133, by using a forward primer targeted to the very beginning of exon 1 of BC019095, and a reverse primer targeted to the very end of exon 13 of AB050197 (Appendix A, Table A.1) (Figure 2.3A).

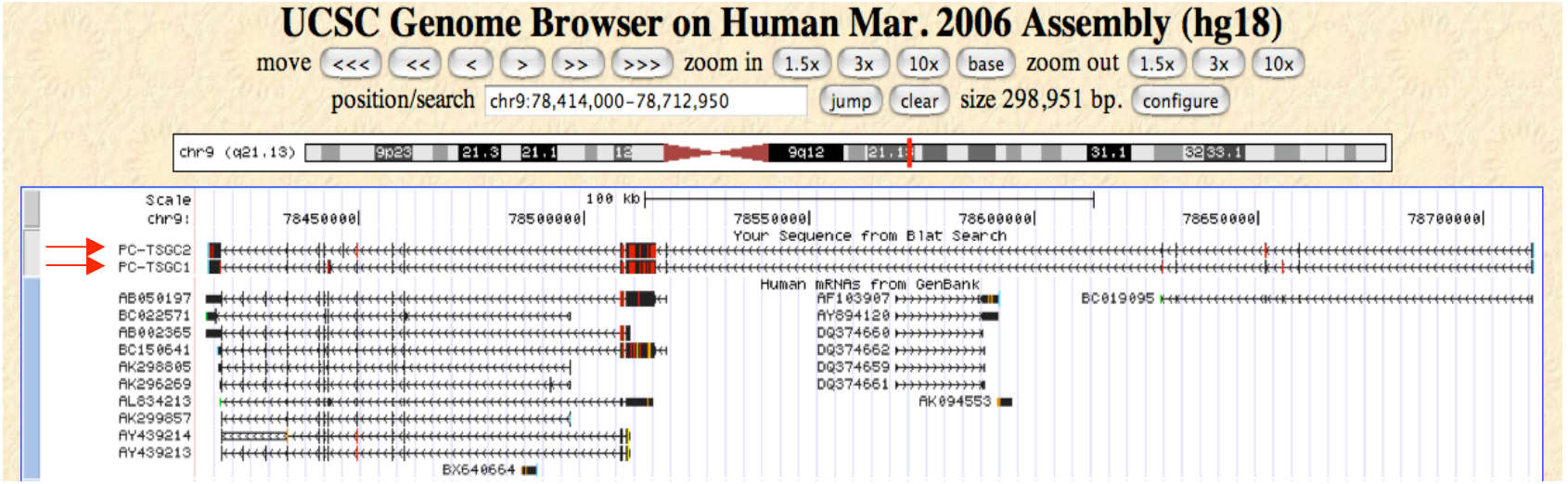
Next, we cloned the 12kb amplicon into a pCR2.1 TOPO-TA vector (Invitrogen). Restriction enzymes analysis and DNA sequencing of the cloned sequences confirmed that we were able to isolate two different isoforms that closely matched our 12 kb predicted sequence (Figures 2.1E and 2.1F). The alignment of both cloned sequences against the human reference genome assembly proved that such sequence is transcribed from a genomic locus harboring a single transcriptional unit that spans almost 300kb on chromosome 9q21 (Figure 2.2).

Based on sequence analysis and functional characterization of the 12 kb cloned sequences (presented in the following chapters), we named this new genetic unit "*PCA3 Controlled – Tumor Suppressor Gene Candidate"* (*PC-TSGC*). The sequences representing two different isoforms of *PC-TSGC* were entered into the GenBank as *PC-TSGC1*

(GenBank accession number: FJ808772, Appendix B.1) and *PC-TSGC2* (GenBank accession number: FJ808773, Appendix B.2)

Figure 2.2

Identification and analysis of full-length PC-TSGC. Alignment of PC-TSGC1 and PC-TSGC2 full-length cloned sequences against the human reference genome assembly on the Genome Browser web server. PC-TSGC1 and PC-TSGC2 are indicated by red arrows.


UCSC Genome Browser on Human Mar. 2006 Assembly (hg18)
move <<< << < > >> >>> zoom in 1.5x 3x 10x base zoom out 1.5x 3x 10x
position/search chr9:78,414,000-78,712,950 jump clear size 298,951 bp. configure
chr9 (q21.13)
Scale
chr9:
100 kb
78450000
78500000
78550000
78600000
78650000
78700000
Your Sequence from Blat Search
PC-TSGC2
PC-TSGC1
Human mRNAs from GenBank
AB050197
BC022571
AB002365
BC150641
AK298805
AK296269
AL834213
AK299857
AY439214
AY439213
BX640664
AF103907
AY894120
DQ374660
DQ374662
DQ374659
DQ374661
AK094553
BC019095

2.3 Sequence analysis of *PC-TSGC1* and *PC-TSGC2*

Both *PC-TSGC1* and *PC-TSGC2* lack two exons at the 3' region of the gene that instead are present in the *BMCC1* transcript; moreover, in the same region, *PC-TSGC2* includes a new exon that was never annotated in any EST. Finally, at the 5' end of the gene, *PC-TSGC2* is alternatively spliced resulting in the exclusion of exon 3 (Figure 2.1E and 2.1F). All these features gain particular significance when an open reading frame (ORF) analysis is performed on both *PC-TSGC1* and *PC-TSGC2*.

PC-TSGC1 is the longer of the two isoform we identified (12,341 nucleotides, Appendix B.1) and has a 9,174 nt open reading frame (ORF). In detail, following a relatively short 5'UTR (nucleotides 1 – 78), this ORF is in frame with a canonical Kozak sequence (GACATGG) (nucleotides 76 - 82) containing the first Methionine start codon (nucleotides 79 – 81). Toward the 3' region of the gene, a TAG stop codon is followed by a large 3' UTR (nucleotides 9,253 – 12,341). *PC-TSGC1* translates for a predicted polypeptide of 3,057 residues.

In the *PC-TSGC2* isoform (12,217 nucleotides, Appendix B.2), alternative splicing of exon 3 causes a frame-shift at the 5' region, while the novel exon at the 3' region introduces an early stop codon. As a consequence, *PC-TSGC2* displays a significantly larger 5' UTR (nucleotides 1 – 409), and it exploits an alternative translation starting site where another canonical Kozak sequence (ACCATGG) (nucleotides 407 –

413) and first ATG codon are found in frame with the main ORF of 8,202 nt. The early stop codon introduced by alternative splicing marks the boundary with an also larger 3'UTR (nucleotides 8612 – 12217). *PC-TSGC2* translates for a predicted truncated polypeptide of 2,733 residues.

Analysis of conserved domains revealed a DHHA2 domain at the amino-termini of both PC-TSGC1 and PC-TSGC2, whereas PC-TSGC1 additionally has an amino-terminal PPX1 domain and a carboxy-terminal BCH/Sec14p-like domain (Figure 2.3B).

The DHHA2 domain stands for DHH Associated domain 2, and is often found adjacent to a DHH domain, which is a conserved domain characterized by a DHH motif (Aravind *et al.,* 1998), however, we did not find any DHH domain in PC-TSGC1. The DHHA2 is diagnostic of DHH subfamily 2 members, which identifies a family of phosphoesterases responsible for the hydrolysis of phosphodiester bonds, including 3',5'-cyclic nucleotides (e.g. cAMP, cGMP) (Middelhaufe *et al.,* 2007). This domain is about 120 residues long and contains a conserved DXK motif at its amino-terminus.

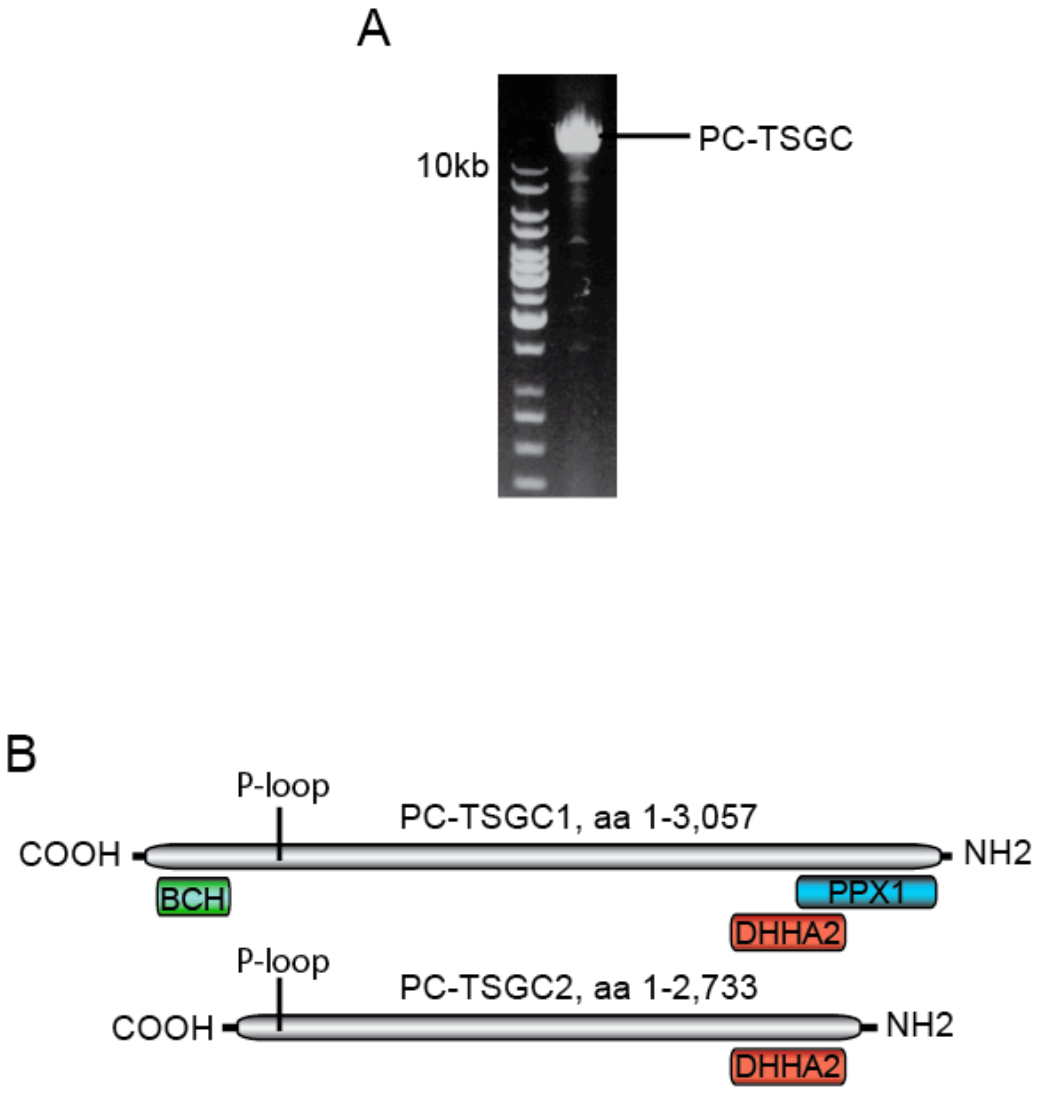
The PPX1 region is an inorganic pyrophosphatase - exopolyphosphatase domain involved in energy production and conversion (Sutera *et al.,* 1999), and the BCH/Sec14p-like region, here referred to as BCH, is generally found in lipid-regulated proteins such as RhoGAPs, RhoGEFs, and neurofibromin (Shang *et al.,* 2003; Zhou *et al.,* 2002). The BCH domain of PC-TSGC1, is similar to the BCH domain

already described for the BMCC1 sequence (Machida *et al.,* 2006), with the only difference of an in-frame 29 amino-acids deletion due to the absence of the two exons at the 3' region of PC-TSGC1, when compared to BMCC1 (Figure 2.1E). Within the BCH domain of PC-TSGC1, a RRRMP motif (residues 2967-2971) corresponds to the RRKMP sequence known to be involved in homophilic/heterophilic interactions between BCH domain-containing proteins, and a conserved RRLRK motif (residues 2985-2989) represents the arginine patch of BCH domains, which is necessary for the activity of a BCH domain as a GTPase Activating Protein (Low *et al.,* 2000). By aligning the two PC-TSGC1 and PC-TSGC2 isoforms, it is clear that not only PC-TSGC2 is truncated at the C-terminus, but also this truncation corresponds exactly with the beginning of the BCH domain of PC-TSGC1 (Figure 2.3B).

Finally, Machida *et al.* (2006) have identified a P-loop motif on the BMCC1 sequence, which we have also found present in both the PC-TSGC isoforms that we cloned (GPGWSGKT, residues 2652-2659 in PC-TSGC1 and residues 2474-2481 in PC-TSGC2). P-loops are usually found in ATP and GTP binding proteins and they have also been described to mediate the activity of pro-apoptotic proteins (Larisch *et al.,* 2000; Saraste *et al.,* 1990).

Figure 2.3

Amplification and sequence analysis of the 12kb PC-TSGC cDNA. *(A*) RT-PCR from the MDA-PCa-133 tumor graft yields the full-length PC-TSGC 12kb amplicon. (*B*) Location of the conserved domains identified on PC-TSGC1 and PC-TSGC2.


A
10kb
PC-TSGC
B
P-loop
PC-TSGC1, aa 1-3,057
COOH
NH2
BCH
PPX1
DHHA2
P-loop
PC-TSGC2, aa 1-2,733
COOH
NH2
DHHA2

CHAPTER 3:

ANALYSIS OF PC-TSGC EXPRESSION AND REGULATION

3.1 PC-TSGC expression in cell lines

We analyzed the expression profile of PC-TSGC in a panel of 22 cell lines derived from several human tissues (Chapter 6: Materials and Methods). We performed quantitative RT-PCR (qRT-PCR) experiments by assaying the expression levels of PC-TSGC at four different regions along the PC-TSGC mRNA (TSGst, TSGmid, TSGsk2, and TSGend) (Appendix A, Table A.1). High levels of expression of PC-TSGC were evident in cell lines derived from brain, breast, and prostate tumors, with the highest in the prostate cancer cell line LNCaP. In contrast, the other two most common prostate cancer cell lines (DU145 and PC3) did not show significant PC-TSGC mRNA levels (Figure 3.1). Thus, from this analysis we selected the LNCaP cell line as the optimal prostate cancer model for the further characterization of PC-TSGC *in vitro*.

Next, we investigated whether PC-TSGC is translated into an actual protein as predicted by our sequence analysis. We assayed LNCaP cell extracts by immunoblot with an anti-PRUNE2 antibody recognizing the amino-terminal region of PC-TSGC: a band of high molecular mass, consistent with the predicted full-length PC-TSGC isoform (337-kDa), was observed (Figure 3.2A).

In order to confirm that the protein detected was actually PC-TSGC, we cloned the PC-TSGC1 and PC-TSGC2 coding sequences into a

pcDNA expression vector (Invitrogen) in frame with a carboxy-terminal V5 tag (Figure 3.2B). We assayed the expression of the recombinant proteins in a negative expression background (from PC3 cell line) by immunoblot and the recombinant PC-TSGC1 band matched the endogenous PC-TSGC band from the LNCaP extract (Figure 3.2A). Notably, PC-TSGC2 was detected only minimally by the anti-PRUNE2 antibody, but this isoform was reactive with the anti-V5 antibody, a result in complete agreement with the amino-terminal truncation of PC-TSGC2, which contains the epitope recognized by the anti-PRUNE2 antibody. This result confirms that PC-TSGC not only is transcribed and processed into a mRNA of 12kb, but also is translated into protein isoforms of 3,057 (PC-TSGC1) and 2,733 (PC-TSGC2) residues, as predicted by our analysis.

Figure 3.1

Expression of PC-TSGC in cell lines. The TSGst, TSGmid, TSGsk2, and TSGend qRT-PCR assays were targeted to four different regions encompassing the entire PC-TSGC full-length sequence (Appendix A, Table A.1). Relative expression levels were compared against a panel of standard endogenous controls (Chapter 6: Materials and Methods).

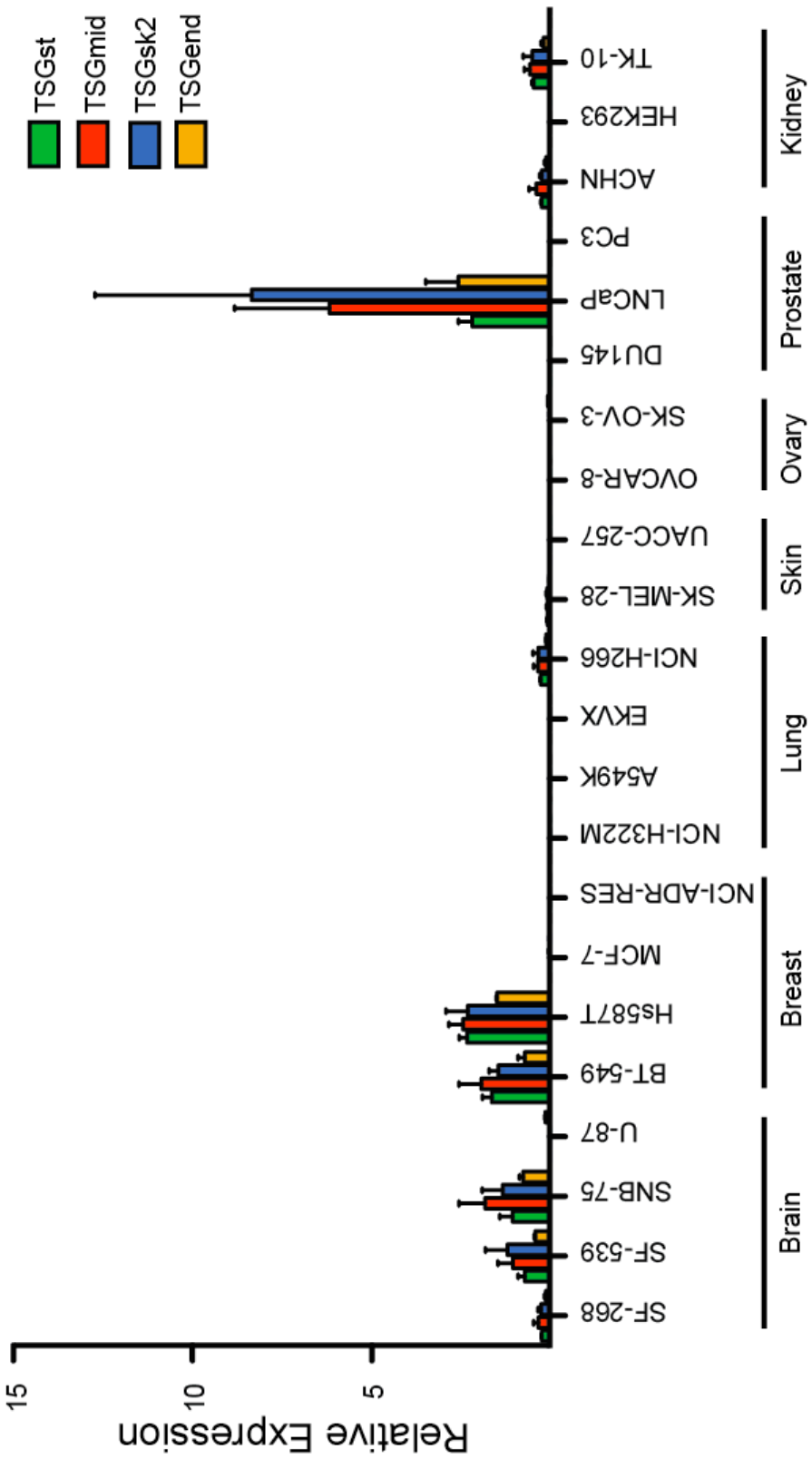
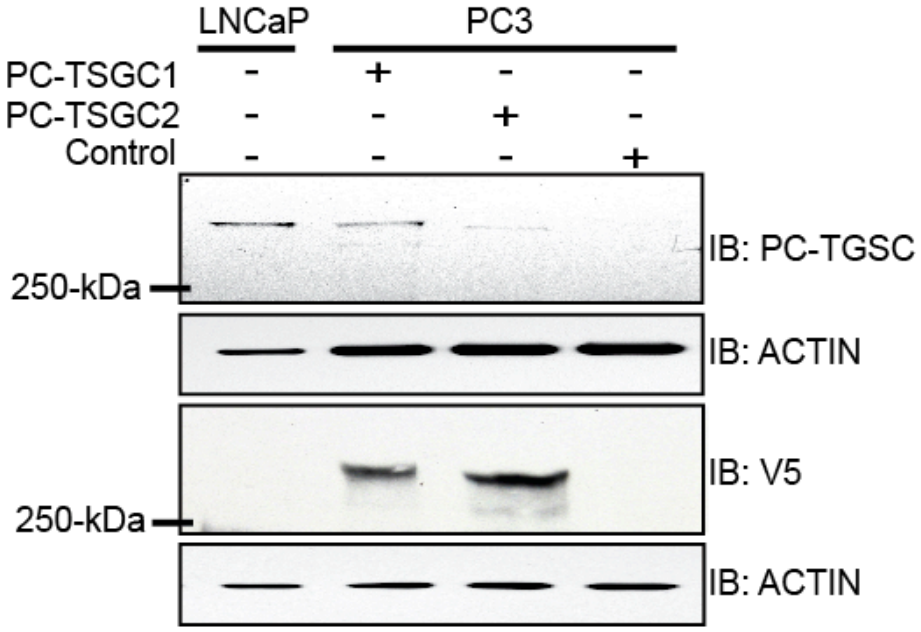

TSGst
TSGmid
TSGsk2
TSGend
Relative Expression
15
10
5
SF-268
SF-539
SNB-75
U-87
BT-549
Hs587T
MCF-7
NCI-ADR-RES
NCI-H322M
A549K
EKVX
NCI-H266
SK-MEL-28
UACC-257
OVCAR-8
SK-OV-3
DU145
LNCaP
PC3
ACHN
HEK293
TK-10
Brain
Breast
Lung
Skin
Ovary
Prostate
Kidney

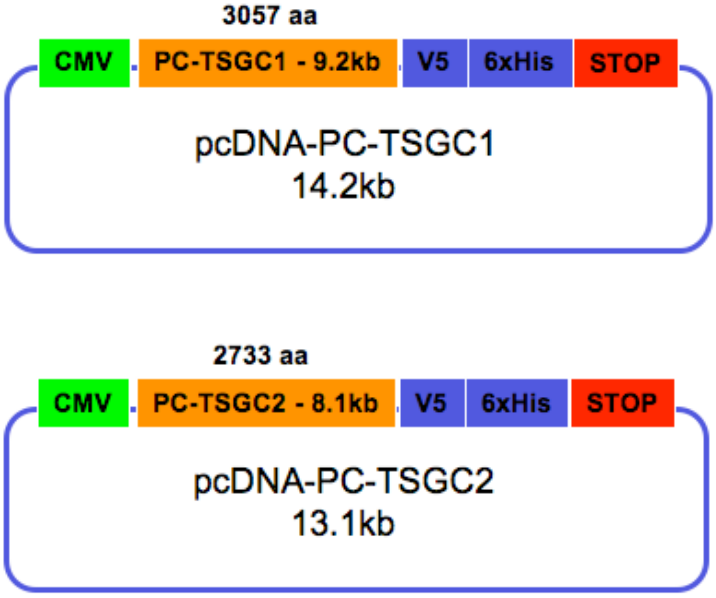
Figure 3.2

PC-TSGC protein expression. (*A*) Full-length PC-TSGC was expressed by LNCaP cells and matched the predicted molecular mass of 337-kDa. Recombinant PC-TSGC isoforms were detected by both anti-PRUNE2 and anti-V5 antibodies in PC3. A GFP expression vector was used as negative control. (*B*) Schematic representation of PC-TSGC1 and PC-TSGC2 expression vectors. Both sequences were cloned into a pcDNA3.1-DEST vector bearing the CMV promoter and V5 and His tags.

A


LNCaP
PC3
PC-TSGC1 - + - -
PC-TSGC2 - - + -
Control - - - +
250-kDa
IB: PC-TGSC
IB: ACTIN
250-kDa
IB: V5
IB: ACTIN

B

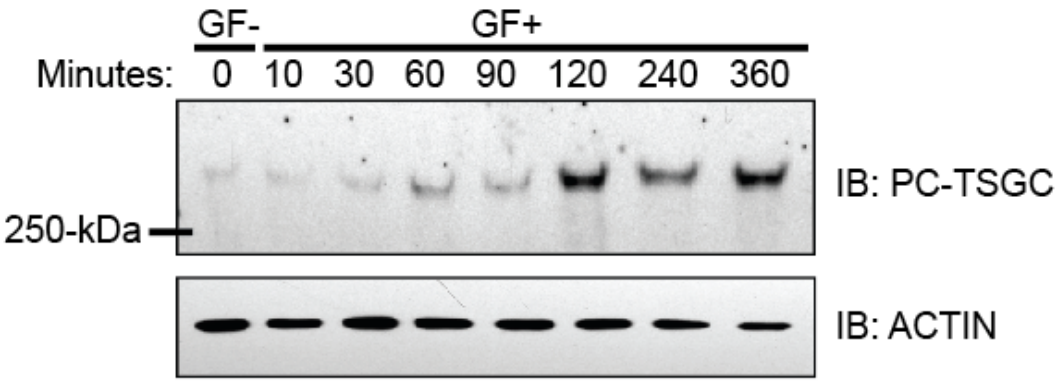

3057 aa
CMV
PC-TSGC1 - 9.2kb
V5
6xHis
STOP
pcDNA-PC-TSGC1
14.2kb
2733 aa
CMV
PC-TSGC2 - 8.1kb
V5
6xHis
STOP
pcDNA-PC-TSGC2
13.1kb

3.2 Induction of PC-TSGC by growth factors stimulation

To begin to understand gene expression and regulation, we asked whether the levels of PC-TSGC were affected by cell culture conditions. We tested LNCaP cells grown under starvation conditions followed by stimulation with a combination of growth factors (GF; Chapter 6: Materials and Methods). We monitored PC-TSGC levels by western blot at different time points and we found that LNCaP cells exhibited increased expression of PC-TSGC over time after stimulation with growth factors (Figure 3.3).

Figure 3.3

Time course analysis of the expression of PC-TSGC in LNCaP cells. Cells are tested under starvation conditions (GF-) and after stimulation with a combination of growth factors (GF+) at the indicated serial time points.


GF-
GF+
Minutes:
0
10
30
60
90
120
240
360
250-kDa
IB: PC-TSGC
IB: ACTIN

3.3 Regulation of PC-TSGC by the ncRNA PCA3

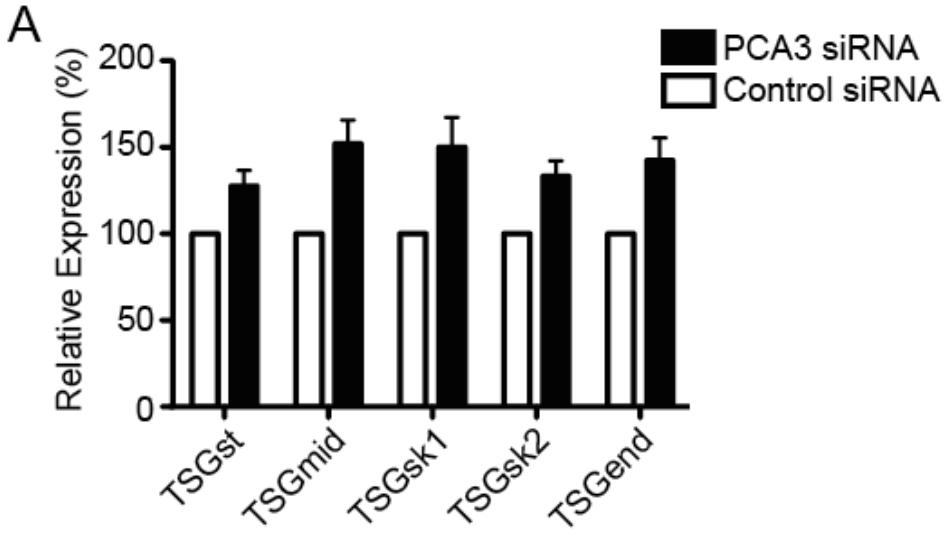
By identifying the full-length PC-TSGC transcript, we also demonstrated that PCA3 is an antisense ncRNA, and precisely, transcribed from an intron of PC-TSGC. Several lines of evidences suggest that intronic-antisense ncRNA can be involved in the regulation of expression levels of the corresponding sense gene (Faghihi *et al.,* 2008; Scheele *et al.,* 2007; Louro *et al.*, 2007; Katayama *et al.,* 2005; Munroe *et al.,* 2006; Tasheva *et al.,* 1995), therefore, we tested whether PCA3 would be able to regulate the expression of PC-TSGC. We specifically down-regulated PCA3 by siRNA in LNCaP cell line and we monitored PC-TSGC expression by qPCR and immunoblot. As a control, we used a validated, non-targeting siRNA under the same experimental conditions.

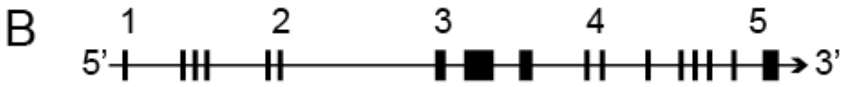
We monitored PC-TSGC expression levels by using five assays targeted to different regions along the PC-TSGC transcript (TSGst, TSGmid, TSGsk1, TSGsk2, and TSGend) (Appendix A, Table A.1). We found that upon reduction of PCA3, all PC-TSGC regions were significantly upregulated. All the assays returned similar levels of up-regulation of the PC-TSGC mRNA (Figures 3.4A and 3.4B), thus by averaging the results from each single assay, we detected a 41% overall increase of PC-TSGC mRNA levels concomitant with a 44% decrease of PCA3 mRNA levels (*t*-test, $p<0.001$; Figure 3.5A).

This result was also confirmed by immunoblot, data indicating that variations in levels of PCA3 can ultimately influence levels of PC-TSGC protein in LNCaP cells (Figure 3.5B).

Figure 3.4

Comparative qRT-PCR analysis of PC-TSGC levels in PCA3 siRNA samples. (*A*) PC-TSGC mRNA levels in the control samples (non-targeted siRNA) were set at 100%; the increased levels of PC-TSGC mRNA are shown as percentage increases tested at five different regions along the PC-TSGC sequence (Appendix A, Table A.1). (*B*) Statistical analysis (Student's *t*-test) for increases in PC-TSGC mRNA upon PCA3 siRNA treatment of LNCaP cells. The numbers mapped to a schematic representation of the PC-TSGC gene indicate the transcript regions tested by qRT-PCR.



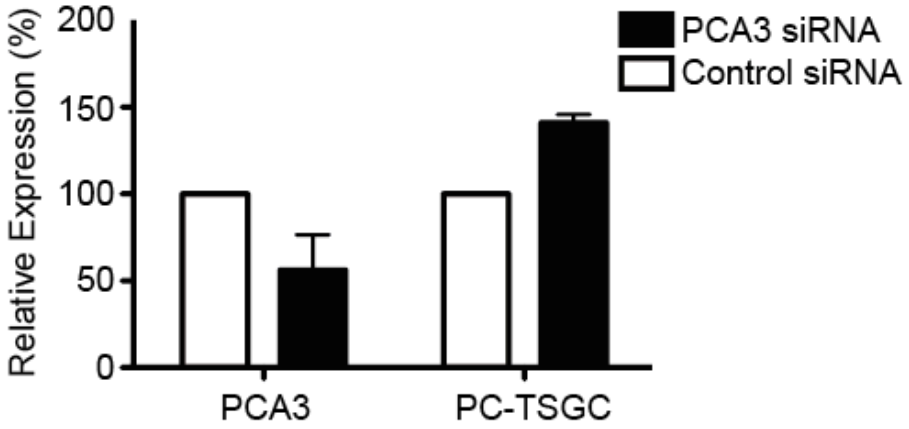


qRT-PCR assay	p-value
1) TSGst	0.038
2) TSGmid	0.019
3) TSGsk1	0.043
4) TSGsk2	0.020
5) TSGend	0.031

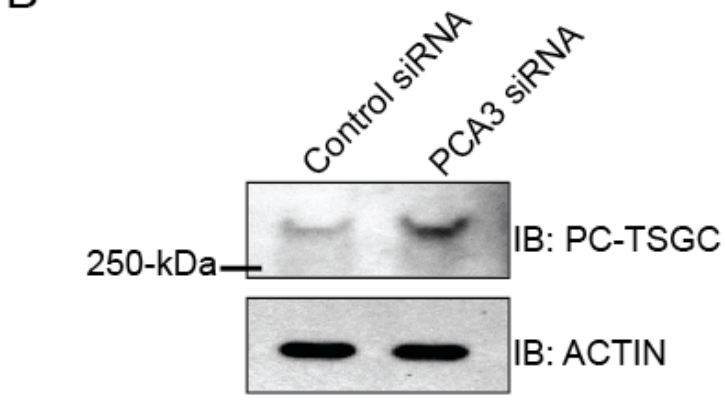
Figure 3.5

Overall variation in PC-TSGC mRNA levels upon treatment of LNCaP cells with PCA3 siRNA. (*A*) PC-TSGC mRNA levels in the control samples (non-targeted siRNA) were set at 100%, and are shown as percentage increases in the PCA3 siRNA sample (*t*-test, $p<0.001$). (*B*) Immunoblot of the variation of PC-TSGC protein levels upon treatment of LNCaP cells with PCA3 siRNA.

A


Relative Expression (%)
200
150
100
50
0
PCA3 siRNA
Control siRNA
PCA3
PC-TSGC

B


Control siRNA
PCA3 siRNA
250-kDa
IB: PC-TSGC
IB: ACTIN

CHAPTER 4:

FUNCTIONAL CHARACTERIZATION OF PC-TSGC

4.1 PC-TSGC associates with both RhoA and nm23-H1

To begin characterizing PC-TSGC, we undertook a sequence analysis-based approach for the identification of possible interacting partners. As already mentioned, we identified three specific conserved domain from the PC-TSGC sequence: the BCH domain at the carboxy-terminus, and the PPX1 and DHHA2 domains at the amino-terminus.

In an attempt to characterize the BCH domain of BMCC1, Soh et al. (2008), were not able to amplify in a single contig the entire BMCC1 sequence, which was originally assembled in vitro from several cDNA fragments (Machida *et al.,* 2006), although they were able to clone a shorter sequence that they named BNIP2XL (BNIP2 Extra Long). BNIP2XL is approximately 750 residues long and comprises the entire carboxy-terminus of BMCC1, including its BCH domain. Sho and coworkers extensively characterized the protein-protein interaction between the BCH domain of BNIP2XL and RhoA in cell lines transfected with two different recombinant isoforms of BNIP2XL (BNIP2XLα, which bears the full BCH domain similarly to BMCC1, and BNIP2XLβ, which bears a BCH domain harboring a 37 amino acids deletion at the carboxy-terminus) and recombinant RhoA, a small GTPase involved in the formation of stress fibers and in the induction of oncogenic transformation. Importantly, they showed that BNIP2XL inhibits RhoA, via its BCH domain (Soh *et al.,* 2008).

Sequence analysis of PC-TSGC shows that the full-length PC-TSGC transcript (PC-TSGC1) includes both BMCC1 and BNIP2XL sequences and that the same BCH domain is found at its carboxy-terminus, with the exception of an in-frame 29 amino acids deletions similar to the one reported for BNIP2XLβ (Figure 4.1A and 4.2A). Thus, it would be reasonable to predict that also the BCH domain of PC-TSGC could associate with RhoA. However, we wanted to test whether such interaction takes place between RhoA and the full-length PC-TSGC protein, and equally importantly, whether the interaction occurs between endogenous proteins.

Endogenous PC-TSGC produced by LNCaP cells was immunoprecipitated and probed for endogenous RhoA. We tested LNCaP cells grown under starvation conditions followed by stimulation with a combination of growth factors (GF). Our results show that PC-TSGC is associated to RhoA upon cell stimulation *in vitro* (Figure 4.2B). The association of PC-TSGC to RhoA upon cell stimulation with growth factors indicates that the observed interaction is specific and inducible.

At the amino terminus instead, the presence of the PPX1 and the DHHA2 domains define a region of the protein that is similar to the protein hPRUNE. hPRUNE is a phosphodiesterase (PDE) belonging to the DHH superfamily of phosphoesterases and pyrophosphatases (D'Angelo *et al.,* 2004). This protein superfamily is characterized by the presence of four distinct motifs, among which the third one is the most conserved one and

contains the DHH sequence, from which the superfamily derives its name (Aravind *et al.,* 1998). Both PC-TSGC and hPRUNE harbor the PPX1 and the DHHA2 domain, however PC-TSGC lacks the DHH motif (Figure 4.2A).

Despite this difference in the domain architecture and the fact that PC-TSGC extends significantly beyond this region, resulting in a polypeptide almost 7-times the size of hPRUNE, we found a high degree of similarity between the entire hPRUNE sequence and the amino terminus of PC-TSGC, including a region of hPRUNE that has been shown to mediate the interaction of hPRUNE with non-metastatic protein 23 (nm23-H1) (Middelhaufe *et al.,* 2007; Reymond *et al.,* 1999) (Figure 4.1B). Therefore, we tested whether PC-TSGC might also associate with nm23-H1. Immunoprecipitation analysis of the endogenous PC-TSGC confirmed that both PC-TSGC and nm23-H1 coimmunoprecipitate within the same complex (Figure 4.2B). Even in this case, we observed that such association is specific and inducible upon cell stimulation with growth factors. This results show that PC-TSGC has at least two protein partners, RhoA and nm23-H1, both of which play roles in the oncogenic transformation and metastatic potential of cancer cells (Coleman *et al.,* 2004, Hartsough *et al.,* 2000).

Figure 4.1

Sequence alignment between PC-TSGC1 and RhoA and nm23-H1. (*A*) Sequence alignment of the BCH domain from BNIP2XLα (full-length domain), BNIP2XLβ (37 amino acid deletion), and PC-TSGC1 (29 amino acid deletion in-frame). The sequences for the BCH domain are highlighted in red. (*B*) Sequence alignment of the carboxy-terminus of hPRUNE (residues 378-453) with the corresponding region in PC-TSGC1 (residues 412-654): the sequence of hPRUNE known to mediate the interaction with nm23-H1 is highlighted in red.

A

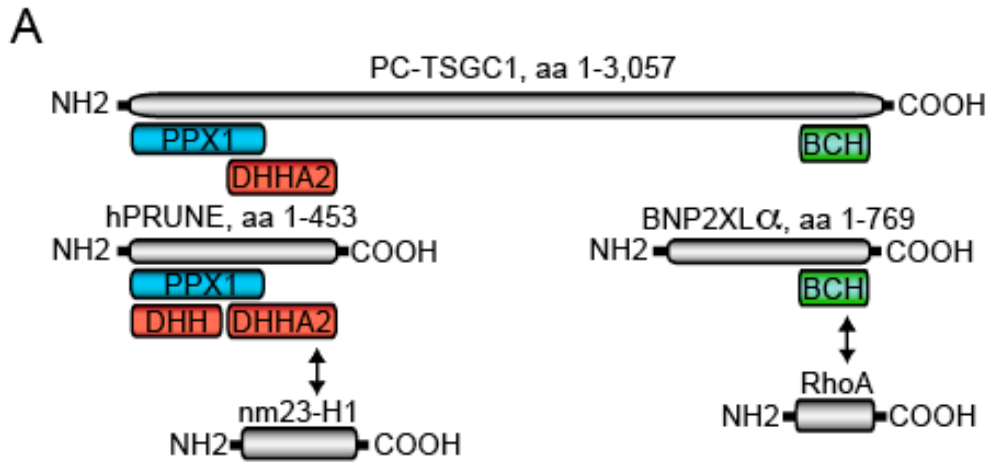
```
BNIP2XLα   VVIGDQEQRIDMKVIEPYRRVISHGGLRGYYGDGLNAIIVFAACFLPDSSRADYHYVMEN 622
BNIP2XLβ   VVIGDQEQRIDMKVIEPYRRVISHGGLRGYYGDGLNAIIVFAACFLPDSSRADYHYVMEN 622
PC-TSGC1   VVIGEQEQRIDMKVIEPYRRVISHGGDSGYYGDGLNAIIVFAACFLPDSSRADYHYVMEN 2939
           ****:**********************  ********************************

BNIP2XLα   LFLYVISTLELMVAEDYMIVYLNGATPRRRMPGLGWMKKCYQMIDRRLRKNLKSFIIVHP 682
BNIP2XLβ   LFLYVISTLELMVAEDYMIVYLNGATPRRRMPGLGWMKKCYQMIDRRLRKNLKSFIIVHP 682
PC-TSGC1   LFLYVISTLELMVAEDYMIVYLNGATPRRRMPGLGWMKKCYQMIDRRLRKNLKSFIIVHP 2999
            ************************************************************

BNIP2XLα   SWFIRTILAVTRPFISSKFSSKIKYVNSLSELSGLIPMDCIHIPESIIKLDEELREASEA 742
BNIP2XLβ   SWFIRTILAVTRPFISSKFSSKIKYVNSLSELSGLIPMDCIHIPESIIKY---------- 732
PC-TSGC1   SWFIRTILAVTRPFISSKFSSKIKYVNSLSELSGLIPMDCIHIPESIIN---------- 3048
           ************************************************:

BNIP2XLα   AKTSCLYNDPEMSSMEKDIDLKLKEKP 769
BNIP2XLβ   --------------------------
PC-TSGC1   ------------------IDLKLKEKP 3057
```

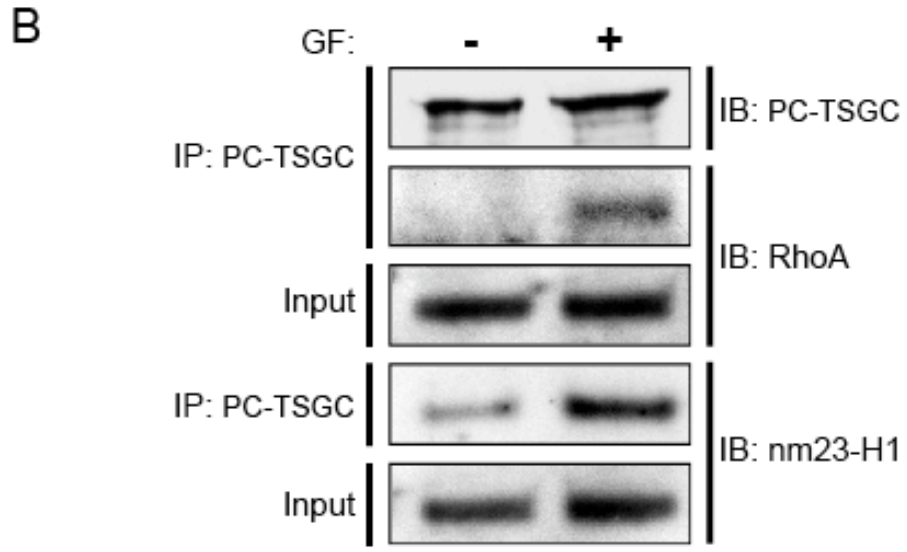
B

```
hPRUNE     VSREQVDKELDRAS---------------------------------------------- 391
PC-TSGC1   SNQAQVDANVDLVSPDSGLATIRSSRSSKESSVFLSDDSPVGEGAGPHHTLLPGLDSYSP 471
            .: *** ::* .*

hPRUNE     ------------------------------------------------------------
PC-TSGC1   IPEGAVAEEHAWSGEHGEHFDLFNFDPAPMASGQSQQSSHSADYSPADDFFPNSDLSEGQ 531


hPRUNE     --------------------NSLISGLSQD------------------------------ 401
PC-TSGC1   LPAGPEGLDGMGTNMSNYSSSSLLSGAGKDSLVEHDEEFVQRQDSPRDNSERNLSLTDFV 591
                             .**:** .:*

hPRUNE     -EEDPP----------LPPTPMNSLVDECPLDQGLPKLSAEAVFEKCSQISLSQSTTASL 450
PC-TSGC1   GDESPSPERLKNTGKRIPPTPMNSLVESSPSTEEPASLYTEDMTQKATDTGHMGPPQTHA 651
            :*.*.          :*********:..*  :  ..* :* : :*.:: .   .. :

hPRUNE     SKK--------------------------------------------------------- 453
PC-TSGC1   RCSSWWGGLEIDSKNIADAWSSSEQESVFQSPESWKEHKPSSIDRRASDSVFQPKSLEFT 711
```

Figure 4.2

Association of PC-TSGC with RhoA and nm23-H1. (*A*) Hypothetical protein interactions with RhoA and nm23-H1: conserved domains are indicated in blue (PPX1), red (DHH and DHHA2), and green (BCH), and protein sequences are indicated in gray. (*B*) Co-immunoprecipitation of PC-TSGC with RhoA and nm23-H1. Lysates from LNCaP wild-type cells under starvation (GF-) or stimulated (10 minutes) with a combination of growth factors (GF+) were tested. The anti-PRUNE2 antibody efficiently immuno-precipitated (IP) endogenous PC-TSGC (IP:PC-TSGC – IB:PC-TSGC); endogenous RhoA associated with PC-TSGC upon growth factors stimulation (IP:PC-TSGC – IB:RhoA); endogenous nm23-H1 associated with PC-TSGC upon growth factors stimulation (IP:PC-TSGC – IB:nm23-H1); total cell extracts prior to IP used as loading controls confirmed that the observed associations were specific and inducible (Input – IB:RhoA; Input – IB:nm23-H1).


A
PC-TSGC1, aa 1-3,057
NH2
COOH
PPX1
DHHA2
BCH
hPRUNE, aa 1-453
NH2
COOH
PPX1
DHH
DHHA2
nm23-H1
NH2
COOH
BNP2XLα, aa 1-769
NH2
COOH
BCH
RhoA
NH2
COOH


B
GF:
-
+
IP: PC-TSGC
IB: PC-TSGC
IB: RhoA
Input
IP: PC-TSGC
IB: nm23-H1
Input

4.2 PC-TSGC participates in the control of anti-apoptotic and cell proliferation signaling pathways

PC-TSGC interacts with RhoA and nm23-H1, and its primary sequence harbors features such as the BCH domain and the P-loop that are found in proteins involved in the regulation of cell proliferation and apoptosis (Low *et al.,* 2000; Larisch *et al.,* 2000).

Because the levels of PC-TSGC increase upon down-regulation of PCA3 (Figures 3.4 and 3.5), we treated LNCaP cells with PCA3 siRNA, and monitored cell viability by using a metabolic assay with WST-1 reagent (Roche). The WST-1 reagent measures the cleavage of a tetrazolium salt to formazan, which directly correlates to the number of metabolically active cells in the culture. Notably, we found that LNCaP cells treated with PCA3 siRNA were ~20% less viable than control cells (*t*-test, $p<0.005$; Figure 4.3), indicating a possible role for PC-TSGC in cell proliferation and/or apoptosis. To confirm this observation, we specifically down-regulated PC-TSGC in LNCaP by shRNA, and subsequently evaluated cell viability (Figure 4.4A). Consistently, we found that PC-TSGC-shRNA cells were ~20% more viable than control cells, both under starvation conditions (*t*-test, $p<0.0001$) and upon stimulation with a combination of growth factors (*t*-test, $p<0.001$) (Figure 4.4B).

Figure 4.3

WST-1 metabolic assay on LNCaP cells treated with PCA3 siRNA. Cells are tested under four different conditions *in vitro*, with increasing concentrations of FBS: the highest reduction in cell viability is observed for cells cultured in media containing 2.5% FBS (~20% reduction; *t*-test, $p<0.005$).

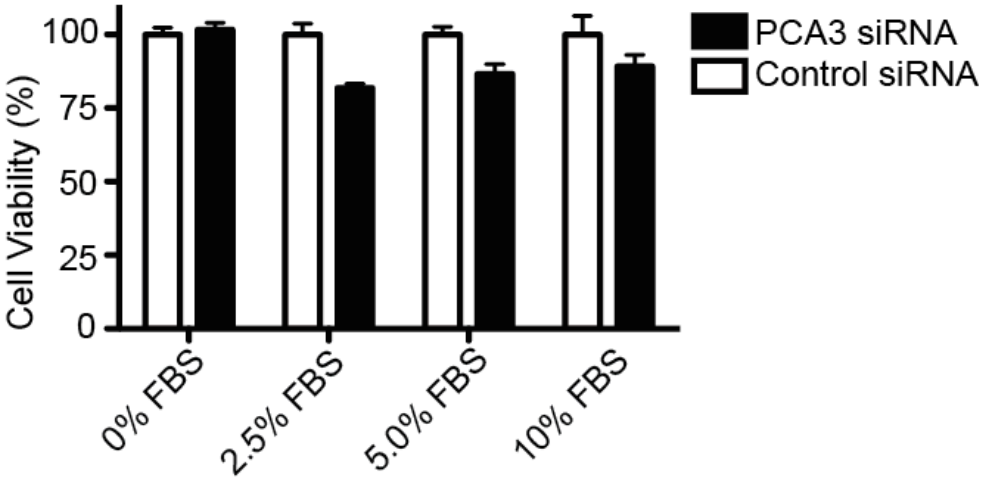
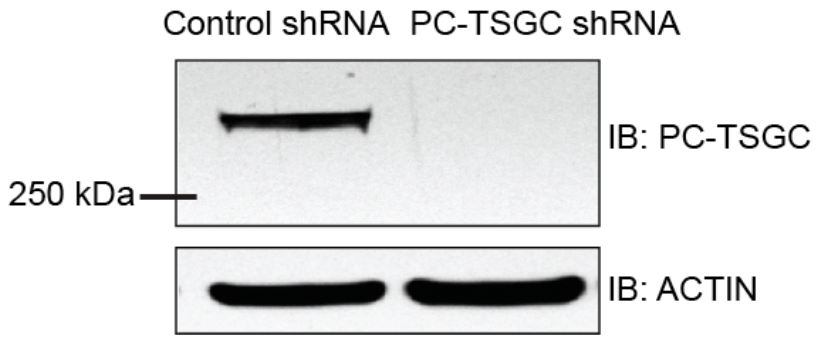

PCA3 siRNA
Control siRNA
Cell Viability (%)
100
75
50
25
0
0% FBS
2.5% FBS
5.0% FBS
10% FBS

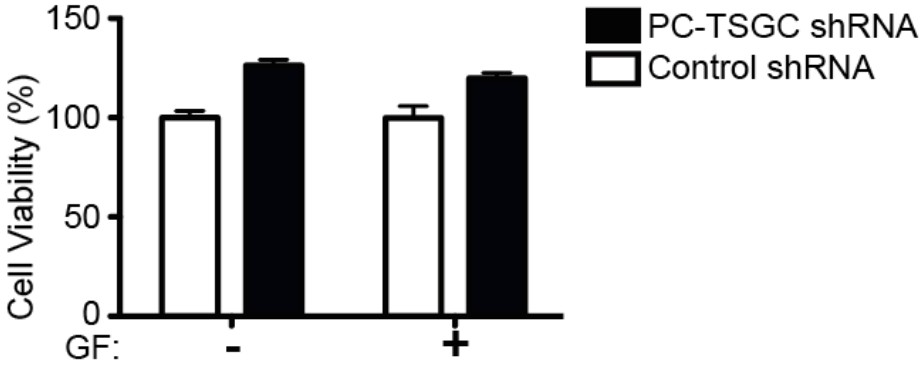
Figure 4.4

Evaluation of the effect of PC-TSGC on cell viability. (*A*) Stable knock-down of PC-TSGC in LNCaP by puromycin selection of PC-TSGC-shRNA cells. (*B*) WST-1 metabolic assay on PC-TSGC-shRNA cells compared with non-targeted control shRNA: differences were tested under starvation conditions (GF-; *t*-test, $p<0.0001$) and after stimulation with a combination of growth factors (GF+; *t*-test, $p<0.001$).

A


Control shRNA
PC-TSGC shRNA
IB: PC-TSGC
250 kDa
IB: ACTIN

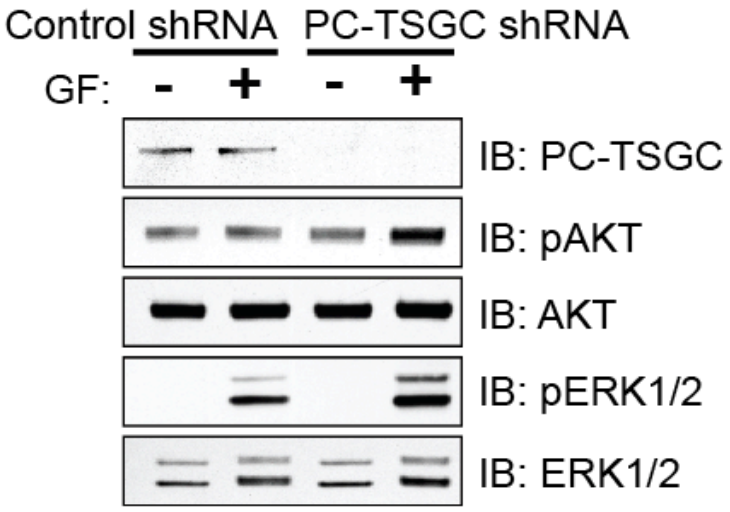
B


150
100
50
0
Cell Viability (%)
GF:
-
+
PC-TSGC shRNA
Control shRNA

Next, we analyzed the phosphorylation status of AKT and ERK1/2, which are involved in the signal transduction of both anti-apoptosis and pro-proliferation pathways. Consistently with the cell viability assays results, we observed that both AKT and ERK1/2 were more activated in PC-TSGC-shRNA cells upon cell stimulation. Notably, the increased level of AKT phosphorylation was still appreciable in PC-TSGC-shRNA cells even if LNCaP cells are known to have high constitutive levels of AKT activation (Nesterov *et al.,* 2001). Likewise, ERK1/2 phosphorylation status was remarkably higher in PC-TSGC-shRNA cells than control cells upon cell induction with growth factors (Figure 4.5). These results indicate that PC-TSGC is likely involved in the control of both anti-apoptotic and pro-proliferative signaling pathways.

Figure 4.5

Effects of PC-TSGC on cell signaling transduction. Induction of ERK1/2 (pERK1/2: phosphorylated ERK1/2) and AKT (pAKT: phosphorylated AKT) signaling pathways by stimulation (10 minutes) with combination of growth factors (GF) in PC-TSGC-shRNA and control shRNA cells.


Control shRNA
PC-TSGC shRNA
GF: - + - +
IB: PC-TSGC
IB: pAKT
IB: AKT
IB: pERK1/2
IB: ERK1/2

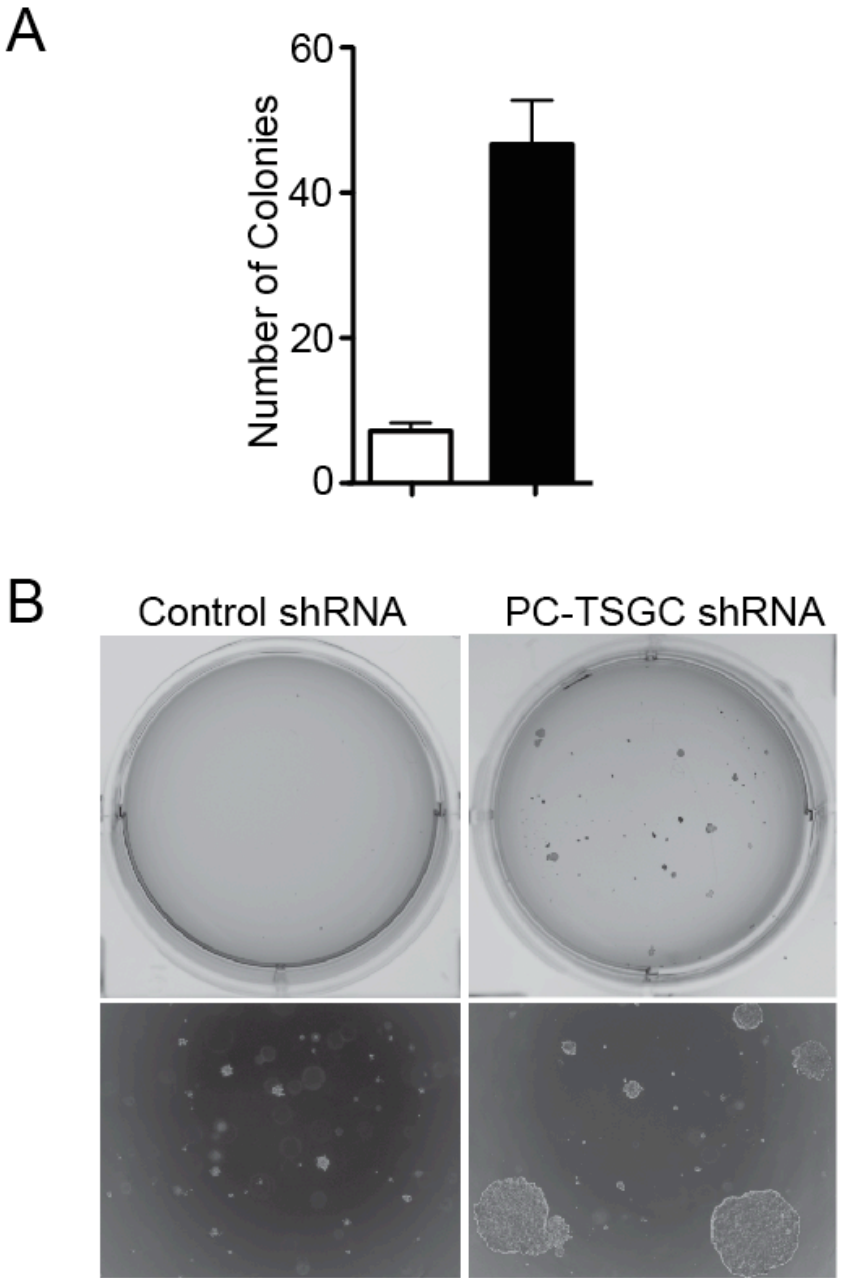
4.3 PC-TSGC prevents anchorage-independent cell growth

To investigate whether PC-TSGC could play a role in the oncogenic transformation of prostate cancer cells, we performed a colony formation assay in semi-solid media (soft agar). The soft agar assay tests the ability of cells to grow without being attached to a solid surface (anchorage-independency) and is one of the most stringent assays for cellular transformation *in vitro*. In fact, the phenotypic changes associated to the acquisition of an anchorage-independent phenotype closely resemble the same changes that occur during the neoplastic transformation of a cancer cell.

With this experiment, PC-TSGC-shRNA cells grew extremely well in soft agar compared to the control (Figure 4.6A and 4.6B; *t*-test, $p<0.0001$), data indicating that a lack of PC-TSGC not only results in increased proliferation but also in anchorage-independent cell growth.

Figure 4.6

Evaluation of the effect of PC-TSGC on anchorage-independent cell growth. (*A*) Colony count of PC-TSGC-shRNA cells and control cells following crystal violet staining. (*B*) Representative image of anchorage-independent cell growth after PC-TSGC down-regulation.


A
60
40
20
0
Number of Colonies
B
Control shRNA
PC-TSGC shRNA

4.4 PC-TSGC prevents tumor growth *in vivo*.

To conclude our analysis, we finally tested whether PC-TSGC effect would also affect tumor growth in tumor bearing mice. We injected PC-TSGC-shRNA cells subcutaneously into SCID mice and then monitored tumor growth at up to three weeks in comparison with mice injected with control-shRNA cells.

LNCaP cells have a very low efficiency rate of tumor formation *in vivo*. Remarkably, in our experiment we did not observe any tumor formation in the mice injected with control cells.

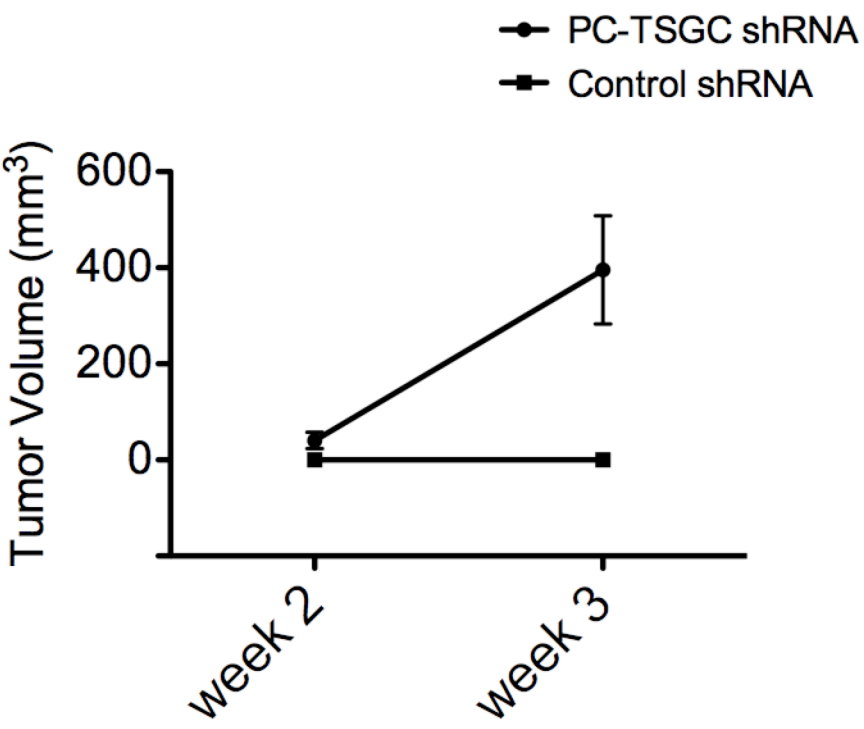
Instead, mice injected with PC-TSGC-shRNA cells exhibited a striking difference in tumor growth, as 100% of mice formed tumors (*t*-test, $p<0.04$) (Figure 4.7).

We demonstrate that a decrease in PC-TSGC levels leads to an increased activation of AKT and ERK1/2 signaling pathways, and the decrease in PC-TSGC confers an anchorage-independent phenotype and favors tumor formation and growth *in vivo.*

Taken together, our results provide evidence that PC-TSGC has a role in determining the transformation and tumor growth potential of prostate cancer cells, and may be considered as a new tumor suppressor gene candidate.

Figure 4.7

Evaluation of the effect of PC-TSGC on tumor growth *in vivo*. The tumor volumes of SCID mice injected subcutaneously with 5 million cells in a 50% volume of Matrigel were monitored at 2 and 3 weeks. PC-TSGC-shRNA injected mice present significant tumor growth compared to control-shRNA injected mice (*t*-test, $p<0.04$).


PC-TSGC shRNA
Control shRNA
Tumor Volume (mm3)
600
400
200
0
week 2
week 3

CHAPTER 5:

DISCUSSION

5.1 Identification of a new human gene: *PC-TSGC*

Until now, PRUNE2 and BMCC1 were two distinct sequences that were annotated as independent genes on the opposite strand and in an antisense orientation to PCA3. Recent studies postulate the existence of a single transcript putatively comprising both PRUNE2 and BMCC1 and extending for approximately 300 kb (Clarke *et al.,* 2009; Salagierski *et al.,* 2009). Such transcript was also predicted to be processed into a 12 kb mRNA, however, these studies were based on partial cDNA fragments, bioinformatics, and *in silico* sequence assembly (GenBank accession number: NM015225), without further genetic or functional data. With our work instead, we provide evidence for a new transcriptional unit on chromosome 9q21, which spans almost 300kb and includes a significant intronic region in correspondence to PCA3. This finding not only provides evidence for a previously unrecognized genetic element, but also at the same time allows classifying PCA3 as an intronic-antisense ncRNA.

We named the newly discovered gene “PCA3 Controlled-Tumor Suppressor Gene Candidate” (*PC-TSGC*) and we identified two splicing isoforms, PC-TSGC1 and PC-TSGC2, which encode polypeptides of 3,057 and 2,733 amino acids, respectively.

We have shown that the expression of PC-TSGC is responsive to stimulation of cells with growth factors in a time-dependent manner. Given the negative effect of PC-TSGC on the viability of prostate cancer cells,

we hypothesize that after stimulation with growth factors, the cells express PC-TSGC, which in turn opposes cell growth and survival, thereby resulting in a negative-feedback loop on the activity of growth signaling pathways.

5.2 Biological role of PCA3 in Prostate Cancer

We have also shown that levels of PC-TSGC are influenced by its intronic-antisense ncRNA PCA3. A decade of research on PCA3 has led to the establishment of this ncRNA as a *bona fide* prostate cancer biomarker; in fact, several diagnostic assays that rely on the detection of PCA3 in biological fluids have been developed (van Gills *et al.,* 2007; de la Taille *et al.,* 2007; Groskopf *et al.,* 2006). Although the importance of testing for PCA3 in human prostate cancer-related applications is widely available (www.pca3.org), a function for this ncRNA and its role in the biology of prostate cancer has not as yet been described. Here, we have shown (i) that levels of PC-TSGC increase upon down-regulation of PCA3, and (ii) that PCA3 down-regulation is associated with a decrease in cell viability, which could be a consequence of the concomitant increase in PC-TSGC. In mammalian cells, only a handful of examples of gene regulation by ncRNA have been reported to date, and mechanistic insights of their function remain largely unknown (Faghihi *et al.,* 2009). PCA3 is a fully intronic, antisense, and spliced ncRNA, which is perhaps one of the least characterized classes of long ncRNA. Our siRNA data demonstrate that PCA3 regulation over PC-TSGC is specifically mediated by the actual PCA3 transcript, rather than occurring via sense-antisense transcriptional interference.

Considering the possible molecular mechanisms of gene expression regulation by antisense long ncRNA, it is very unlikely that PCA3 as a ncRNA could elicit chromatin modifications or promoter methylation that would in turn shut down the expression of PC-TSGC. In fact, these kind of epigenetic regulations are usually mediated by antisense ncRNAs that are located in the proximity of the promoter of the corresponding sense gene (Faghihi *et al.,* 2008; Scheele *et al.,* 2007; Munroe *et al.,* 2006). Furthermore, any DNA methylation or chromatin remodeling would drastically impair the transcription of both PC-TSGC as well as PCA3, which is a condition that is not compatible with our experimental observations. Therefore, because we show that PCA3 and PC-TSGC are simultaneously expressed at different levels in LNCaP cells, and because we demonstrate an inverse correlation between the RNA level of both genes, we hypothesize that the PCA3 regulation over PC-TSGC likely depends on the formation of a PCA3:PC-TSGC nuclear RNA duplex, which may affects the transport, and/or stability of PRUNE2 primary transcripts in prostate cancer cells.

PCA3 is mainly localized in the nucleus (Schalken *et al.,* 2003) but most importantly, is fully intronic to PC-TSGC. Because PCA3 does not overlap with any intron-exon junction of PC-TSGC, we would exclude a potential regulation of PCA3 of the alternative splicing of PC-TSGC. Instead, a likely regulatory function of PCA3 could involve an RNA editing-mediated mechanism that would affect PC-TSGC stability and/or nuclear

localization. This mechanisms would depend on the recognition of the PCA3:PC-TSGC RNA nuclear duplex by the RNA editing enzyme ADAR1 (Keegan *et al.,* 2004). According to the degree of RNA editing, the PC-TSGC RNA would then face different fates, including sequestration in the nucleus and/or subsequent degradation.

PCA3 presents multiple transcriptional start sites as well as polyadenylation sites, and the complexity of its transcriptome renders the identification of a precise mechanism of action a challenging task. However, to our knowledge, this report is the first demonstration of a biological function for the ncRNA PCA3 towards the regulation of PC-TSGC and cellular viability in prostate cancer.

5.3 Biological role of PC-TSGC in Prostate Cancer

We have also shown that sequence analysis and functional assays in human LNCaP cells confirm that PC-TSGC affects cell viability, and that a decrease in PC-TSGC results in anchorage-independent cell growth and tumor formation *in vivo*. These represent possible mechanisms potentially linked to the protein interaction of PRUNE2 with RhoA and nm23-H1.

A detailed work of characterization performed by Soh and coworkers on the BCH domain of BNIP2XL demonstrate that this domain is able to interact and inhibit RhoA in human cell lines. This process includes the sequestration of RhoA and the inhibition of its binding to the Lbc-RhoGEF, which is the specific guanine exchange factor (GEF) responsible for RhoA activation (Soh *et al.,* 2008). We demonstrated that endogenous PC-TSGC and RhoA specifically interact upon cell stimulation by growth factors. Considering that BNIP2XL is almost identical to the caboxy-terminus of the full-length PC-TSGC protein, and that a lack of PC-TSGC results in increased cell proliferation, it can be assumed that PC-TSGC may exert its function at least in part by preventing the binding of RhoA to its activator Lbc-RhoGEF. However, since the BCH domain has been shown to possess a GTPase activating protein (RhoGAP) function (Shang *et al.,* 2003), we cannot exclude that

PRUNE2 could also directly promote hydrolysis of GTP by the GTP-bound RhoA, directly terminating any downstream RhoA signaling.

RhoA signaling induces, via distinct pathways, the phosphorylation of both ERK1/2 and AKT (Del Re *et al.,* 2008; Basile *et al.,* 2007; Coleman *et al.,* 2004). In particular, the expression of RhoA stimulates cell cycle progression and proliferation by inducing the G1/S phase transition in quiescent cells (Olson *et al.,* 1995). Several mechanisms have been suggested for such function, including the repression of CDK inhibitors such as p21 and p27, or the induction of cyclinE/CDK2, and in general, an active RhoA signaling is required for a sustained activation of the ERK signaling pathway, which eventually leads to the progression through the G1 to the S phase (Coleman *et al.,* 2004). At the same time, RhoA modulates cell fate and survival by engaging a signaling pathway through focal adhesion kinase (FAK). Activation of FAK leads to activation of PI3K and concomitant phosphorylation of AKT (Del Re *et al.,* 2008).

Our experiments with shRNA show that cells that lack PC-TSGC exhibit enhanced rates of cell viability, a condition confirmed by an increased activation of both ERK1/2 and AKT signaling pathways. Given the RhoA-inhibitory properties of the BCH domain (Soh *et al.,* 2008), it is possible that decreased levels of PC-TSGC lead to more sustained RhoA signaling, therefore resulting in increased cell viability.

On the other hand, we found common conserved domains between the amino-terminus of PC-TSGC and the protein hPRUNE, which is able to interact and inhibit nm23-H1 in breast cancer. We demonstrated that PC-TSGC is also able to interact with nm23-H1 upon stimulation with growth factors, however the interaction of PC-TSGC with nm23-H1 leaves more open questions.

hPRUNE is the human homologous of the Drosophila PRUNE protein which was first described for the brownish eye color phenotype of PRUNE mutant flies (Galasso *et al., 2009;* Lifschytz *et al.,* 1969). The Drosophila PRUNE protein is known to interact with the awd (abnormal wing disc) gene family, whose human homologous proteins are known as the nm23 gene family (Timmons *et al.,* 1995). Nine different proteins have been described for this family, coding for proteins with anti-metastatic properties. In humans, hPRUNE is able to interact and to inhibit nm23-H1 and this inhibition results in increased motility and metastatic potential for breast cancer cells (D'Angelo *et al.,* 2004). nm23-H1 is well-known for preventing metastasis formation in certain type of cancers, however it is not yet clear how exactly nm23-H1 exerts this function (Galasso *et al.,* 2009). The nm23 proteins have several known biochemical activities: as nucleoside diphosphate kinase (NDPK), as phosphotransferase and histidine protein kinase, as a 3'-5'-exonuclease, and as a regulator of GTP binding proteins by direct interaction with guanine exchanging factors (GEF) activators. However none of these functions has been shown to be

essential and directly correlated to nm23 anti-metastatic activity. In addition, low levels of nm23-H1 have generally been correlated with more aggressive tumors, but only in certain types of cancers (e.g., breast, melanoma, sarcoma and colon), whereas for others (e.g., neuroblastoma and Hodgkin lymphoma) the opposite correlation has been observed (Hartsough *et al.,* 2000). The biological significance of this complex expression profile is not understood; particularly in regard to prostate cancer, the ultimate effect of the expression of nm23-H1 in the prostatic tumor is still controversial.

For instance, several reports have associated the expression of nm23-H1 in prostate cancer with both anti-metastatic and anti-proliferative properties, whereas others have linked it to an increased malignant potential and aggressiveness of prostate cancer cells (Kim *et al.,* 2003; Lee *et al.,* 1999; Igawa *et al.,* 1994; Konishi *et al.,* 1993). In this scenario, the interaction of PRUNE2 with nm23-H1 raises several challenging questions due to the controversy surrounding the multiple biological functions of nm23-H1 and its role in different type of cancer.

In this context, several nm23 protein family members interact with and inhibit guanine-exchange factor (GEF) activators for GTP-binding proteins. In particular, nm23-H1 inhibits Tiam1, a Rac-specific GEF (Miyamoto *et al.,* 2009; Otsuki *et al.,* 2001). One might indeed speculate that, through the RhoA-inhibitory action of its BCH domain, and the recruitment of nm23-H1 that in turn inhibits a specific Rac activator,

PRUNE2 could perhaps orchestrate a combined action towards the down-regulation of different Rho-GTPase family members (RhoA and Rac), thus preventing oncogenic transformation of prostate cells as well as inhibiting tumor growth (Figure 5.1).

The working model proposed in Figure 5.2 summarizes the findings presented in this thesis, and proposes a novel mechanism of action that involves PCA3 and PC-TSGC in the tumorigenesis process of prostate cancer cells.

Figure 5.1

Proposed mechanism of action of PC-TSGC toward the downregulation of signal transduction of Rho GTPase family members. PC-TSGC inhibits the binding of RhoA to its activator Lbc-RhoGEF by direct interaction with RhoA through the BCH domain, and recruits nm23-H1 which in turn inhibits Tiam1, a specific Rac activator. GEF: Guanine nucleotide Exchange Factor; GAP: GTPase Activating Protein.

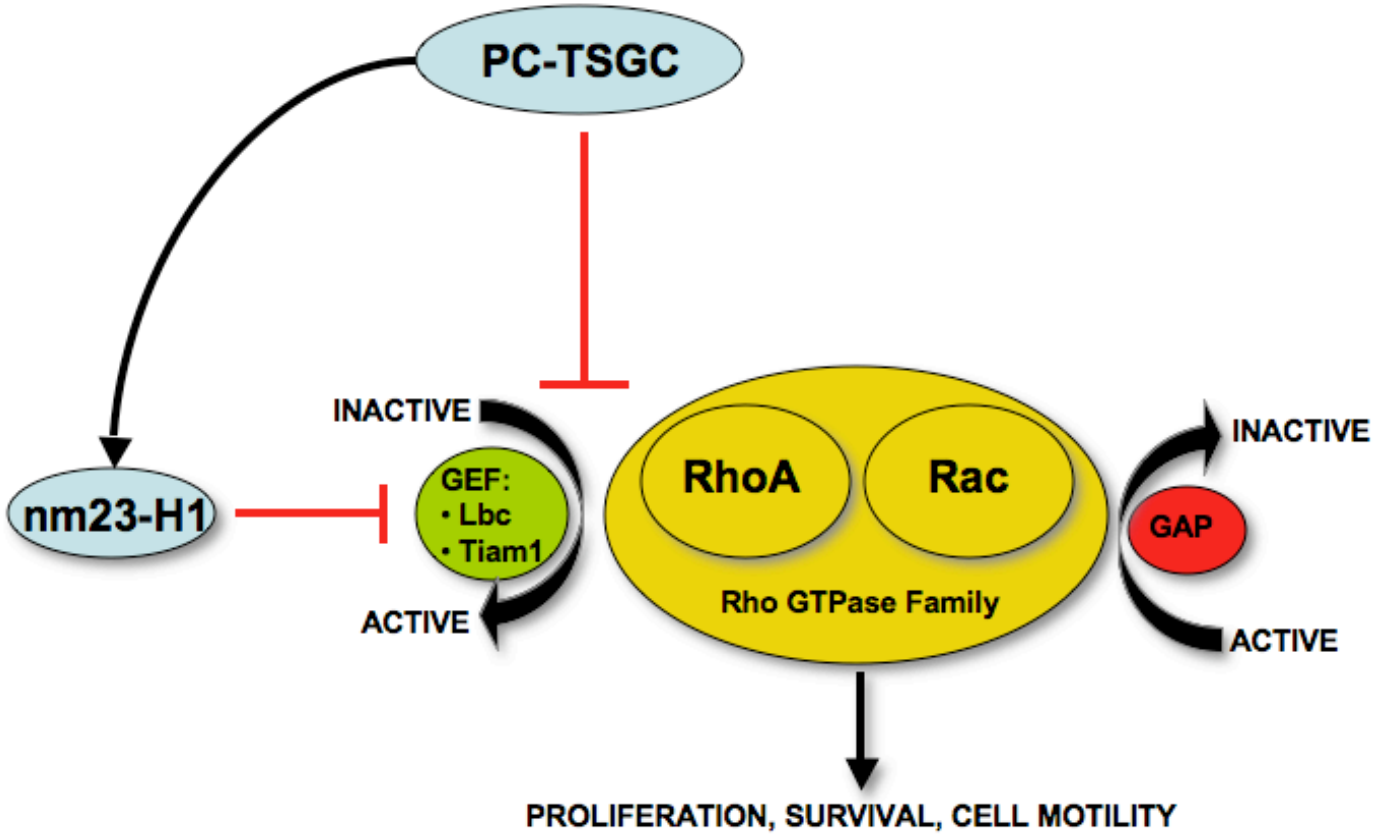
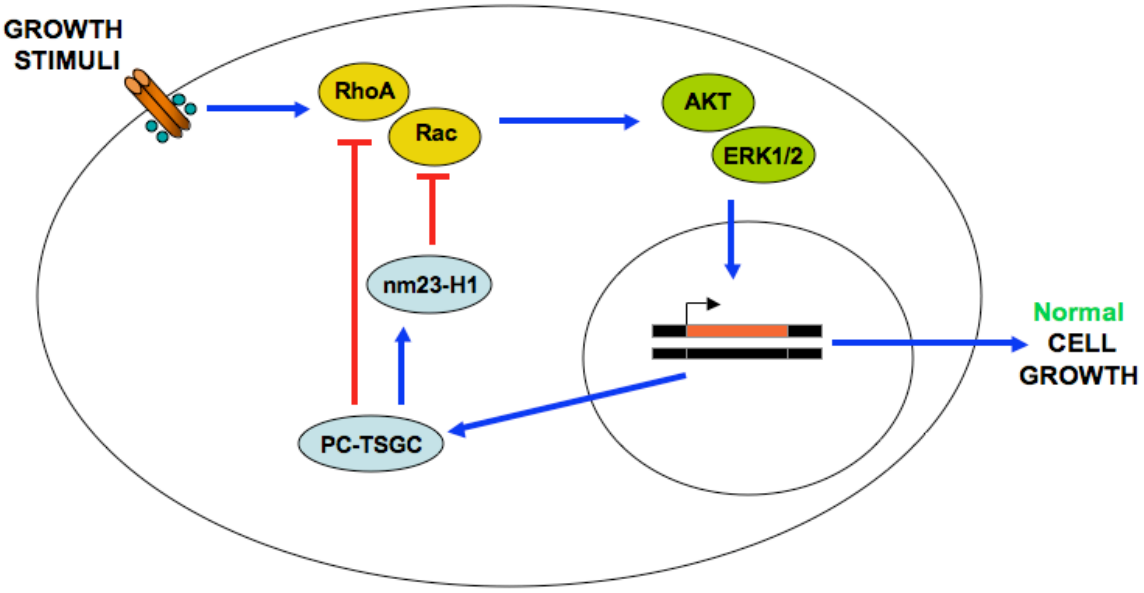

PC-TSGC
nm23-H1
INACTIVE
GEF:
• Lbc
• Tiam1
ACTIVE
RhoA
Rac
Rho GTPase Family
INACTIVE
GAP
ACTIVE
PROLIFERATION, SURVIVAL, CELL MOTILITY

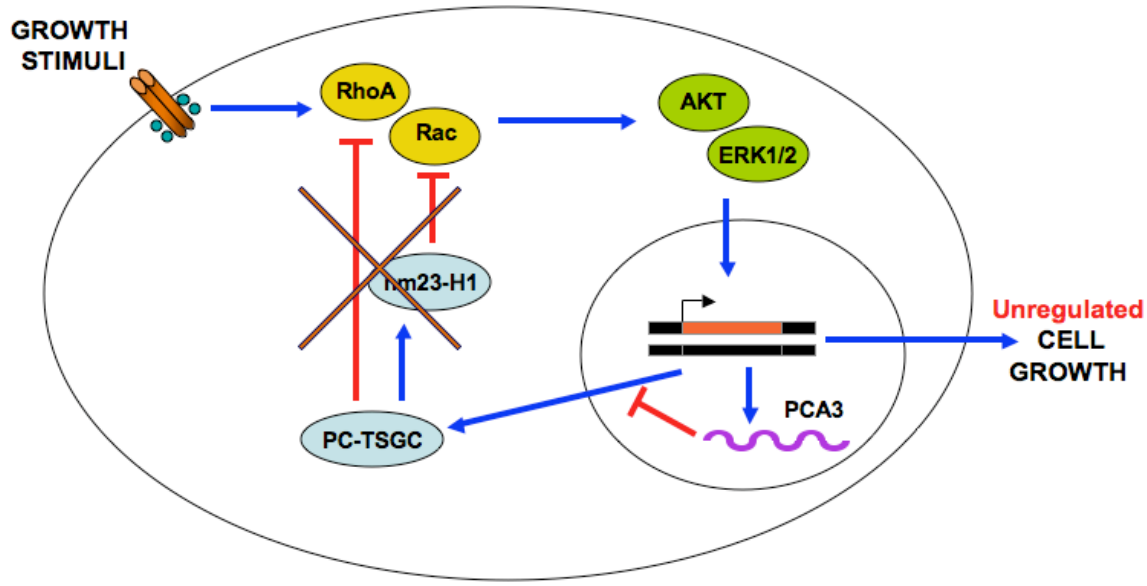
Figure 5.2

Proposed biological roles of PCA3 and PC-TSGC in prostate cancer. (*A*) Normal cell: growth stimuli are signaled to the nucleus through multiple pathways that include activation of RhoA and Rac and subsequent phosphorylation of AKT and ERK1/2. The signal transduction cascade stimulates gene expression in order to initiate cellular replication and inhibit apoptosis. Simultaneously, the same signals elicit the expression of PC-TSGC which in turn inhibits RhoA and Rac (through nm23-H1), thereby resulting in a negative-feedback loop on the activity of cell growth signaling pathways. (*B*) Cancer cell: in a malignant cell, the same mechanism is altered by the abnormal expression of PCA3, which opposes the expression of PC-TSGC. As a result, the control over the RhoA and Rac signaling pathways is lost, and the cell engages an unregulated cell growth that potentially leads to oncogenic transformation.

A


GROWTH
STIMULI
RhoA
Rac
AKT
ERK1/2
nm23-H1
Normal
CELL
GROWTH
PC-TSGC

B


GROWTH
STIMULI
RhoA
Rac
AKT
ERK1/2
nm23-H1
Unregulated
CELL
GROWTH
PCA3
PC-TSGC

5.4 General implications and conclusions

With this work, we identified, cloned, and characterized PC-TSGC as a new tumor suppressor candidate, and we showed that PC-TSGC is regulated by its intronic-antisense ncRNA PCA3. These findings, aside from describing for the first time a new human gene as well as a biological role for the ncRNA PCA3 in prostate cancer, have also more general implications in cancer biology, providing a link between tumor suppression mechanisms and the RNA world.

Cancer is a complex genetic disease that involves the activation of oncogenes and the inactivation of tumor suppressor genes. At the same time, ncRNAs are also involved in several complex genetic diseases, including cancer.

The recognized mechanism of tumor suppression inactivation are based on both genetic and epigenetic mechanisms, including promoter methylations, chromatin remodeling, point mutations and loss of heterozygosity, homozygus deletions, and mono- or bi-allelic mutational inactivations. In addition to these established mechanisms, with the results reported in this thesis we describe the interplay between the ncRNA PCA3 and the tumor suppressor candidate PC-TSGC. Such interplay proposes the existence of another possible mechanism of tumor suppression inactivation that is based on down-regulation mediated by a ncRNA.

The translational and clinical implications of our findings are also relevant. First, PC-TSGC could be explored as a new prostate cancer biomarker. By further analyzing its expression in tumor tissue samples it would be interesting to evaluate any correlation between PC-TSGC expression levels and prostate cancer clinical outcomes. Second, PC-TSGC should be regarded as a possible tool to be employed in the gene therapy of prostate cancer. Even though the size of PC-TSGC would introduce some restrains in the choice of a suitable gene therapy vector for its delivery, the striking tumor suppression potential of PC-TSGC renders this new gene an interesting candidate for such therapeutic approach. Third, we have uncovered new protein-protein interactions that propose new pathways involved in prostate cancer initiation and progression. Such pathways could be identified as new targets for the development of novel therapeutic strategies in the fight against prostate cancer.

In conclusion, this is the first report on the full characterization of PC-TSGC and PCA3 genes, and it identifies PC-TSGC as a new tumor suppressor gene candidate that is down-regulated by a ncRNA in prostate cancer cells. We provide evidence that the regulation of cell growth in the prostate might be subject to fine-tuning that balances the expression and

activity of PCA3 and PC-TSGC to ensure cellular homeostasis. A disruption of this regulation could play a functional role in human prostate cancer.

CHAPTER 6:

MATERIALS AND METHODS

6.1 Bioinformatics and sequence analysis

Chromosomal locations, annotated transcripts, spliced ESTs and sequence mapping were visualized on the Genome Browser web server (genome.uscs.edu). Conserved domain analysis was performed on the NCBI website (www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi), and sequence alignments, with the ClustalW software (www.ebi.ac.uk/clustalw).

6.2 Development of the prostate cancer tumor graft, MDA-PCa-133

MDA PCa 133 tumor graft was developed by the prostate cancer biological models facility (David Koch Ctr. for Applied Research in Genitourinary Cancers) at the University of Texas, MD Anderson Cancer Center. Human tissue specimen used to developed MDA PCa 133 tumor graft was residual from surgery of a bone metastasis of a prostate cancer that was undergoing castrate resistant progression. Small pieces of the human prostate cancer were implanted into subcutaneous pockets of 6- to 8-week-old male CB17 SCID mice (Charles River Laboratories). Tumor developed within 6 months of implantation and was maintained by

passage through mouse as the cells did not sustain in vitro growth. Written informed consent had been obtained from the patient before sample acquisition and sample was processed according to a protocol approved by our institutional review board.

MDA-PCa-133 tumor graft has high expression of Prostate Serum Antigen (PSA).

6.3 Cell lines and cDNA samples

All human cell lines (n=22) of our panel (LNCaP, PC3, DU145, SF-268, SF-539, SNB-75, U-87, BT-549, Hs587T, MCF-7, NCI-ADR-RES, NCI-H322M, A549K, EKVX, NCI-H266, SK-MEL-28, UACC-257, OVCAR-8, SK-OV-3, ACHN, HEK293, TK-10) were cultured in RPMI containing 5% Fetal Bovine Serum (FBS). Total RNAs from cell lines were isolated with the RNeasy mini kit (Qiagen, Hilden, Germany) or the Norgen All-in-One Purification kit (Norgen Biotek Corp, Thorold, ON, Canada). Total RNAs from human tissue samples (prostate, brain, liver, kidney, breast, lung, pancreas, spleen, and testis) were purchased from Stratagene (La Jolla, CA, USA). Total MDA-PCa-133 RNA was obtained from an early passage of the prostate cancer tumor graft MDA-PCa-133. cDNAs were synthesized with Superscript III reverse transcriptase (Invitrogen

Corporation, Carlsbad, CA, USA) from 1µg of total RNA, with N15 random pentadecamers or oligo dT primers.

6.4 DNA Cloning

The sequences of all oligonucleotide primers used are presented (Appendix A, Table 1). Full-length PC-TSGC was amplified from the MDA-PCa-133 cDNA with the KAPA HiFi DNA polymerase (KAPA Biosystems, Cambridge, MA, USA). The isolated band was gel-purified, cloned into TOPO-TA pCR2.1 (Invitrogen), and sequenced. Next, PC-TSGC1 and PC-TSGC2 coding sequences were re-amplified and subcloned into a pENTER/D-TOPO cloning vector (Invitrogen). Finally, each coding sequence was inserted into a pcDNA-DEST40 expression vector in frame with V5 and His tags, by the use of the Gateway recombination technology (Invitrogen). Each cloning step was confirmed by DNA sequencing and restriction enzyme mapping.

6.5 Quantitative RT-PCR

qRT-PCR analysis were performed in a Fast-7500 Real Time PCR system with SYBR-green (Applied Biosystems, Foster City, CA, USA). Gene expression levels were normalized against three different standard endogenous controls (P0 large ribosomal protein (RPLP0), beta-glucuronidase (GUSB), and TATA box-binding protein (TBP)) (Appendix A). Relative expression levels were analyzed according to the $\Delta\Delta C_t$ method (Livak *et al.,* 2001).

6.6 Transfections and Western blot analysis

PC-TSGC1 and PC-TSGC2 were transfected into PC3 cells with the FugeneHD transfection reagent (Roche, Indianapolis, IN, USA). Transgene expression was tested 72 hours after transfection. A GFP-pcDNA3.1 construct was used as a negative control.

Total cell protein extracts were prepared with the Norgen All-in-One Purification kit and were normalized according to the BCA protein assay (Pierce, Rockford, IL, USA). One to 5 µg of protein was loaded onto a 4-12% bis-Tris NuPAGE gel (Invitrogen); proteins were resolved and were subsequently blotted on a nitrocellulose filter (BioRad, Hercules, CA,

USA). PC-TSGC isoforms were detected with an anti-PRUNE2 antibody (ProteinTech Group, Inc, Chicago, IL, USA) or an anti-V5 antibody (ABcam, Cambridge, MA, USA). Antibodies against total-AKT, phospho-AKT, total ERK1/2, and phospho-ERK1/2 (Cell Signaling Technology, Beverly, MA, USA) were used to analyze signaling function. Endogenous levels of actin were monitored with a horseradish peroxidase-conjugated anti-actin antibody (Santa Cruz Biotechnology, Santa Cruz, CA, USA) as a loading control for each blot.

6.7 Cell induction

Subconfluent LNCaP cells were starved for 24 h in RPMI containing 0.25% bovine serum albumine (BSA), and then induced with a combination of growth factors to final working concentrations of VEGF (10 ng/ml), FGF (10 ng/ml), EGF (10 ng/ml), IGF (10 ng/ml), and supplemented with Heparin (5 units/ml).

6.8 Small interfering RNA and short hairpin RNA

A custom-ordered Silencer Select siRNA against PCA3 (5'-GGTTATACTTACTAGCACA-3') (Ambion, Austin, TX, USA) was transfected with the NeoFX transfection reagent (Ambion) into LNCaP cells at a final concentration of 10nM. PCA3 knock-down was assayed by qRT-PCR 72 hours after transfection. For stable PC-TSGC knock-down, a pLKO.1 lentiviral vector expressing a shRNA against PC-TSGC (5'-GCTCAGGGTTTAAGTATTGAA-CTCGAG-TTCAATACTTAAACCCTGAGCT-3') (Open Biosystems, Huntsville, AL, USA) was amplified and co-transfected into 293FT cells with Lipofectamine 2000 (Invitrogen), together with a third generation lentiviral packaging mix. Lentiviruses were produced and used to infect sub-confluent LNCaP cells. Stable clones were maintained under puromycin selection (10µg/ml). Validated non-targeting siRNA (Ambion) or shRNA (Open Biosystem) were used as negative controls.

6.9 Co-immunoprecipitation

Subconfluent LNCaP wild-type cells grown in 10 cm dishes were starved for 24 hours and were subsequently induced for 10 minutes with

the combination of growth factors. Cells were washed with ice-cold PBS and lysed in 1% NP-40, 50 mM Tris-HCl, 135 mM NaCl, 1 mM EDTA, 1 mM β-glycerol-phosphate supplemented with anti-protease (Complete-Mini tablets, Roche) and anti-phosphatase (PhoSTOP tablets, Roche) mixes. Extracts were normalized with the BCA protein assay (Pierce) and were incubated overnight with 2µg of anti-PRUNE2 antibody (ProteinTech Group, Inc) containing 1% BSA. Lysates were incubated with protein A-agarose (4 hours) and the immunocomplexes were washed and resolved on a 4-12% bis-Tris NuPAGE gel (Invitrogen). Blotted filters were probed with anti-PRUNE2, anti-RhoA (Abcam), and anti-nm23 (Abcam) antibodies. Total cell extracts collected from each lysate prior to the immunoprecipitation step were also resolved on the same gel, probed, and used as loading controls (input).

6.10 Cell viability assay

Starved LNCaP cells were transfected in serum-free media with PCA3 siRNA oligos. Serial concentrations of FBS were added after 6 h. Cell viability was measured 72 h later by WST-1 cell viability reagent (Roche). PC-TSGC-shRNA LNCaP cells (or control) were starved for 24 h,

and then stimulated with the combination of growth factors. Cell viability was measured by WST-1 72 h later.

6.11 Soft agar assay

PC-TSGC-shRNA LNCaP cells and control cells were counted and resuspended in 1.5 ml of 0.35% agarose in growth medium, and seeded in 6-wells dishes (5000 cells/well) previously filled with 1ml of 0.7% basal agarose in growth medium. Another layer of 0.7% agarose in growth medium was then added and cells were fed once per week. After 3 weeks, pictures were taken with an inverted microscope at 4x magnification, and colonies were stained with 0.005% crystal violet, and counted.

6.12 Tumor formation assay

PC-TSGC-shRNA LNCaP cells and control cells were counted and resuspended to a final concentration of 5 million cells in 100µl of PBS. Cells were then mixed in 50% volume of Matrigel and a total of 200µl of resuspended cells were then injected subcutaneously in the flank of 7-

weeks old male SCID mice (Charles River). Tumor growth was monitored weekly, and tumor volumes were measured at two and three weeks from injection.

APPENDIX A

Table A.1

List of primers used for qRT-PCR analysis and cloning of full-length PC-TSGC isoforms.

Primer name	Task	Sequence (5' → 3')
TSG.fw	Full-length PC-TSGC amplification	ACCCCGCTCGTCTTCCTT
TSG.rv		CCAAAACGAAGTCTAACAGACA
TSG1enter.fw	PC-TSGC1 and PC-TSGC2 coding sequences sub-cloning	CACCATGGAAGAATTTTTGCAACG
TSG1enter.rv		AGGCTTTTCTTTCAGCTTCAAGTC
TSG2enter.fw		CACCATGGAATCAGAGAAGATCTCAG
TSG2enter.rv		GAGAAGGTTACCTGAATCTCCTCC
TSGst.fw	PC-TSGC qPCR assays	CCACGACATGGAAGAATTTTTG
TSGst.rv		GCGTTTGCTTCGATTCAGTTT
TSGmid.fw		GGAGACCCAGTTCAGTGCTC
TSGmid.rv		TGTAAATGCTTTCAAGTCACTGGT
TSGsk1.fw		CGTTTATTTGCCGGTAGGAG
TSGsk1.rv		GCTCAGGCTCTTTGGTAGGA
TSGsk2.fw		GGGAAATGCTTTCACCACAG
TSGsk2.rv		CTCTTCAAAGGGGATGTCCA
TSGend.fw		TCAATAGCTTATCAGAACTCAGTGG
TSGend.rv		TCAACAGAACCATGAACCAGA

Primer name	Task	Sequence (5' → 3')
RPLP0.fw	Endogenous controls qPCR assays	CGCTGCTGAACATGCTCAAC
RPLP0.rv		ATTGTCGAACACCTGCTGGAT
GUSB.fw		TGATCGCTCACACCAAATCC
GUSB.rv		CCTTGTCTGCTGCATAGTTAGAGTTG
TBP.fw		CCACAGTGAATCTTGGTTGTAAACTT
TBP.rv		AAACCGCTTGGGATTATATTCG

APPENDIX B

B.1 PC-TSGC1 (GenBank accession number: FJ808772)

Total bp: 12,341

5’ UTR: 78 bp (1 – 78)

Kozak sequence: 76 – 82

ATG: 79 – 81

Coding Sequence (in blue): 9174 bp (79- 9252)

Stop codon: 9250-9252

3’ UTR: 3089 bp (9253 – 12341)

Full-length cDNA sequence:

ACCCCGCTCGTCTTCCTTGGGCCGAGATTTTCCACTGCGCCCCTCCGAGTACCC
GGGTTCCAAACCCCTAGCCACGAC**ATG**GAAGAATTTTTGCAACGCGCCAAATCT
AAACTGAATCGAAGCAAACGCTTGGAGAAGGTCCATGTGGTTATTGGGCCTAAA
TCGTGTGACTTGGATTCTCTCATTTCTACCTTCACATATGCTTACTTTCTAGAC
AAGGTCAGTCCACCAGGGGTTCTGTGTTTACCAGTGCTGAACATACCAAGAACT
GAATTCAACTACTTCACCGAGACGAGGTTTATTTTAGAAGAGCTAAATATTTCC
GAATCATTCCACATATTCCGGGATGGAATTAACCTGCATCAGCTAAATGATGAA
GGGAAGTTATCGATAACACTTGTTGGCAGCAGTGTGCTGGCGAGTGAAGACAAA
ACTTTAGAATCAGCAGTTGTCAAAGTCATTAATCCGGTTGAGCAGAGCGATGCC
AACGTTGAGTTCCGAGAGTCTTCCTCTTCTCTCGTGCTAAAGGAGATTCTCCAA
GAGGCTCCTGAGCTCATCACCGAGCAACTGGCTCATCGCCTCAGAGGTAGCATT
CTTTTCAAGTGGATGACCATGGAATCAGAGAAGATCTCAGAGAAGCAGGAGGAA
ATTCTTTCTATCCTGGAAGAAAAATTTCCTAACTTGCCTCCAAGAGAGGACATC
ATCAACGTCCTACAGGAGACCCAGTTCAGTGCTCAGGGTTTAAGTATTGAACAG
ACAATGTTGAAAGATCTAAAGGAGCTGTCAGATGGAGAAATAAAAGTGGCCATT
AGTACTGTGAACATGAACCTTGAGAATTGTCTATTTCACAGCAATATTACCAGT
GACTTGAAAGCATTTACAGACAAGTTTGGTTTTGATGTCCTCATCCTGTTCTCC
AGCTATCTGTCAGAGGAGCAGCAGCCGAGACGACAGATTGCTGTGTACTCAGAA
AACATGGAGCTGTGCAGTCAGATTTGCTGTGAGCTGGAAGAGTGTCAGAACCCT
TGCCTAGAACTGGAGCCCTTTGACTGTGGCTGTGATGAGATCCTGGTGTACCAA
CAAGAGGACCCTTCAGTGACTTGTGATCAGGTGGTTCTCGTTGTCAAGGAAGTC
ATCAACAGGAGGTGTCCAGAGATGGTCTCCAATAGCCGGACATCCTCAACAGAA
GCCGTGGCAGGCAGTGCCCCCCTCTCCCAGGGGTCTTCTGGGATTATGGAATTG
TATGGTTCTGACATAGAGCCACAACCCAGCTCTGTGAATTTCATAGAGAACCCT
CCAGATCTCAATGATTCTAACCAGGCTCAGGTGGATGCCAATGTAGACCTTGTT

AGCCCAGACAGCGGACTGGCTACCATTAGGAGCAGCCGCTCATCCAAGGAGAGC
TCTGTTTTCCTCAGTGACGACAGCCCCGTGGGAGAAGGTGCTGGGCCTCACCAC
ACCCTTCTCCCAGGGCTTGACTCCTACAGCCCCATCCCTGAAGGGGCGGTAGCG
GAGGAACATGCATGGTCTGGAGAACACGGTGAGCACTTCGACCTCTTCAATTTT
GACCCAGCACCCATGGCTTCTGGGCAGTCCCAGCAATCTTCTCATTCTGCAGAC
TACTCCCCAGCAGATGACTTCTTCCCCAACAGCGACCTGTCAGAAGGACAGCTC
CCCGCTGGGCCTGAAGGACTTGATGGCATGGGAACCAACATGTCTAATTATTCA
TCCAGTTCACTTTTGTCAGGGGCTGGCAAAGATAGCCTTGTGGAACATGATGAG
GAGTTTGTCCAGAGACAAGACAGTCCCAGAGATAACTCTGAAAGAAATTTGAGC
CTGACAGATTTTGTGGGAGATGAATCCCCTTCCCCAGAAAGGCTAAAAAATACT
GGAAAGAGGATCCCACCAACACCCATGAATAGTTTAGTAGAAAGCTCGCCATCC
ACTGAAGAACCAGCCTCACTCTATACCGAAGATATGACCCAAAAAGCAACTGAC
ACAGGTCACATGGGGCCACCTCAGACCCATGCACGGTGCAGCAGCTGGTGGGGT
GGTTTGGAAATTGACTCCAAAAATATTGCAGATGCGTGGAGTTCCAGTGAACAG
GAATCTGTTTTCCAGAGCCCTGAATCATGGAAAGAGCATAAGCCAAGCTCCATT
GATAGGAGAGCCTCAGATTCTGTATTTCAACCAAAGAGCCTCGAATTTACAAAG
TCAGGTCCCTGGGAGTCTGAATTTGGTCAGCCTGAACTGGGTAGCAATGATATT
CAAGACAAAAATGAGGAAAGCTTGCCGTTCCAGAACCTGCCCATAGAGAAGTCA
CCTTTGCCAAATACATCTCCCCAAGGAACAAACCACCTGATAGAAGACTTTGCT
TCTTTGTGGCATTCTGGTCGCTCTCCCACAGCCATGCCCGAGCCCTGGGGAAAT
CCTACAGATGATGGTGAACCAGCAGCTGTGGCGCCATTCCCAGCCTGGAGTGCA
TTTGGTAAAGAAGATCATGATGAAGCTTTAAAAAATACCTGGAATTTGCACCCA
ACAAGCAGCAAGACACCTTCTGTTAGGGACCCGAATGAGTGGGCCATGGCAAAA
AGTGGGTTTGCCTTTTCTTCTTCAGAACTACTGGACAATTCACCCAGTGAGATA
AACAATGAAGCAGCTCCAGAAATCTGGGGCAAGAAAAACAATGACTCCAGGGAT
CACATCTTTGCACCTGGAAATCCCAGTTCTGATCTGGATCACACATGGACTAAT
TCTAAGCCACCAAAAGAAGATCAGAATGGTTTAGTGGATCCTAAAACTAGGGGC
AAGGTATATGAAAAGGTAGACTCCTGGAACCTTTTTGAGGAGAATATGAAGAAA
GGAGGGTCAGATGTCCTAGTTCCTTGGGAAGATTCCTTCTTATCTTACAAATGT
TCTGATTACAGTGCATCCAACCTAGGAGAAGATTCGGTGCCTTCCCCCTTAGAT
ACCAATTATTCCACCTCAGACTCTTACACATCACCAACATTTGCTGGAGACGAA
AAGGAAACTGAACACAAGCCATTTGCTAAAGAGGAAGGTTTTGAGTCAAAAGAT
GGTAACTCCACGGCAGAGGAGACTGACATTCCTCCTCAGTCACTGCAACAGTCA
TCTCGAAATCGAATCAGTTCAGGTCCTGGGAACCTAGACATGTGGGCTTCACCT
CATACAGATAACAGTTCTGAAATAAATACCACTCACAACCTGGATGAAAATGAA
CTCAAGACAGAGCACACAGATGGTAAGAATATCTCCATGGAGGATGACGTCGGG
GAAAGCAGCCAGTCCAGTTACGACGACCCCAGCATGATGCAACTGTACAATGAA
ACAAACCGACAACTCACACTTTTGCACAGCAGCACCAACTCCCGGCAGACGGCC
CCTGACAGTCTCGACTTGTGGAACAGAGTGATTTTGGAGGATACTCAGTCCACT
GCAACGATCTCAGATATGGACAATGATTTGGACTGGGATGACTGCAGTGGGGGT
GCGGCAATCCCCAGTGATGGTCAAACAGAAGGATACATGGCTGAAGGTTCCGAG
CCGGAAACCCGATTTACAGTGAGACAGCTGGAACCCTGGGGCTTGGAGTATCAG
GAAGCAAATCAGGTAGATTGGGAGCTCCCTGCCTCTGATGAGCATACCAAGGAC
AGTGCTCCCAGTGAACATCACACATTGAATGAGAAAAGTGGGCAGCTAATTGCA
AACAGTATTTGGGATTCCGTCATGAGAGATAAGGACATGTCATCATTCATGTTA
CCAGGCTCCTCACATATCACAGATTCAGAGCAAAGGGAATTGCCTCCTGAAATC
CCCAGCCATTCAGCAAATGTTAAAGACACTCATTTCCCAGATGCGCCAGCAGCC
TCTGGAACCAGTGAATCAGAGGCACTTATATCTCATCTTGACAAGCAGGACACA

GAGAGGGAAACCCTGCAAAGTGATGCAGCATCCTTGGCGACTAGGCTTGAGAAT
CCAGGGTATTTTCCACACCCAGATCCATGGAAAGGTCATGGCGATGGACAAAGT
GAAAGTGAGAAGGAAGCCCAGGGAGCCACTGACAGGGGGCACCTTGATGAAGAG
GAGGTGATCGCCTCTGGTGTGGAGAATGCCTCAGGGATTTCTGAAAAAGGGCAG
AGTGACCAGGAACTGTCTTCTCTGGTTGCATCTGAACATCAGGAAATCTGCATT
AAATCAGGCAAAATCAGCTCTCTTGCTGTCACTTTCAGTCCTCAAACCGAGGAA
CCAGAGGAAGTTTTAGAGTATGAGGAGGGGTCTTACAATCTAGACTCCCGTGAT
GTGCAAACAGGGATGTCCGCAGATAACCTGCAGCCAAAAGATACCCATGAAAAA
CACCTCATGAGTCAAAGAAATTCAGGTGAAACTACTGAGACTTCAGATGGGATG
AATTTCACAAAATATGTATCTGTACCTGAAAAGGATCTTGAGAAAACTGAAGAA
TGTAACTTTCTAGAGCCAGAGAACGTGGGTGGAGGGCCACCTCACAGAGTTCCC
CGAAGTCTTGACTTTGGGGACGTCCCTATAGACAGTGATGTGCATGTCAGCAGC
ACACGTTCTGAGATAACCAAAAATCTTGACGTTAAGGGGTCTGAAAATAGCCTT
CCAGGAGCCGGTTCGTCTGGAAATTTTGACAGAGATACTATTTCTAGTGAGTAT
ACTCATTCAAGTGCATCAAGTCCTGAGTTAAATGACTCTTCAGTTGCACTGTCC
TCCTGGGGCCAGCAACCCAGTTCTGGGTATCAAGAAGAAAACCAAGGCAACTGG
AGTGAACAGAATCACCAAGAATCTGAACTAATTACCACTGATGGCCAAGTAGAA
ATAGTTACCAAAGTGAAGGATTTAGAGAAAAACAGAATAAATGAGTTTGAAAAG
AGCTTTGATCGCAAAACTCCTACATTTTTAGAGATCTGGAATGACTCAGTTGAT
GGTGATTCCTTTTCCTCTTTATCCAGTCCTGAAACAGGCAAATATTCTGAACAT
TCAGGGACACATCAGGAAAGCAATCTAATTGCTAGCTACCAGGAGAAAAATGAA
CATGACATTTCTGCAACTGTGCAGCCAGAGGATGCCAGGGTCATTTCAACAAGC
TCAGGTTCTGATGATGACAGTGTCGGTGGTGAAGAGTCAATAGAGGAAGAGATC
CAGGTGGCCAACTGCCACGTTGCTGAGGATGAATCCAGAGCTTGGGATTCATTG
AATGAATCTAATAAGTTCTTGGTCACAGCTGATCCTAAGTCTGAAAATATTTAT
GACTACCTAGACAGCTCAGAGCCAGCAGAGAATGAGAATAAGTCAAACCCATTC
TGTGACAATCAACAAAGCAGCCCTGATCCCTGGACTTTCTCACCATTAACGGAG
ACTGAAATGCAGATTACAGCAGTGGAGAAGGAGAAGAGATCTTCTCCAGAAACA
GGGACAACAGGAGATGTTGCATGGCAAATATCTCCCAAAGCTTCGTTCCCAAAG
AACGAAGATAATTCTCAACTGGAAATGCTGGGCTTCTCAGCTGATAGCACTGAG
TGGTGGAAGGCCTCACCCCAGGAAGGGAGACTAATTGAAAGTCCATTTGAAAGG
GAGCTGTCTGACTCCAGTGGTGTGTTGGAGATAAATTCTTCAGTACACCAAAAT
GCCAGTCCCTGGGGAGTACCAGTTCAGGGTGATATTGAGCCCGTGGAAACACAC
TATACTAATCCTTTTAGTGACAACCATCAGTCACCCTTTCTGGAAGGTAATGGG
AAGAACTCCCATGAGCAACTCTGGAACATTCAACCAAGGCAGCCAGACCCAGAT
GCTGACAAGTTCAGCCAGCTTGTAAAATTAGACCAAATTAAAGAAAAGGACTCA
AGAGAGCAAACCTTTGTGTCTGCTGCTGGTGATGAGCTGACTCCTGAAACACCT
ACCCAAGAGCAGTGTCAGGACACCATGCTGCCAGTCTGTGATCATCCGGACGCA
GCCTTTACTCACGCAGAGGAAAATAGTTGTGTTACATCTAATGTTTCAACTAAT
GAAGGTCAAGAAACAAATCAGTGGGAACAAGAAAAATCATACCTAGGTGAGATG
ACAAATTCAAGCATTGCCACAGAAAATTTTCCTGCTGTCAGTTCTCCCACCCAA
CTGATAATGAAGCCAGGCTCTGAATGGGATGGCTCTACCCCAAGTGAGGACTCC
CGAGGTACCTTTGTGCCAGATATTTTACATGGCAACTTTCAAGAGGGTGGGCAG
CTGGCCTCTGCCGCGCCTGACTTGTGGATAGATGCTAAGAAGCCCTTCAGTTTG
AAAGCAGATGGTGAGAATCCTGATATCCTGACGCACTGCGAACATGACAGCAAT
TCTCAGGCTTCCGACAGCCCTGATATATGTCACGATTCTGAAGCAAAGGAAGAG
ACTGAGAAGCACCTCAGTGCTTGCATGGGACCTGAAGTGGAATCCAGTGAGCTT
TGTCTCACTGAGCCAGAGATAGATGAAGAACCCATTTATGAGCCTGGACGGGAG

TTTGTCCCATCCAATGCAGAACTCGATTCTGAAAACGCAACTGTGCTGCCTCCA
ATTGGCTATCAAGCAGACATAAAGGGCTCCTCTCAGCCCACCTCTCATAAAGGT
TCACCTGAACCTTCTGAAATAAACGGTGACAACAGTACAGGTTTACAAGTATCA
GAAAAAGGAGCCAGCCCAGATATGGCACCAATTTTGGAACCAGTTGACAGAAGA
ATCCCAAGGATTGAAAATGTGGCAACTAGCATTTTTGTAACTCACCAAGAGCCA
ACTCCAGAAGGTGACGGTTCTTGGATATCAGACAGCTTTTCTCCTGAAAGTCAG
CCTGGTGCAAGAGCTTTGTTTGATGGTGATCCACATTTATCCACAGAGAATCCT
GCCTTGGTTCCTGATGCTTTGCTAGCCTCAGACACTTGTCTGGATATAAGCGAA
GCTGCCTTTGACCACAGTTTCAGCGATGCCTCAGGTCTCAACACATCCACGGGA
ACAATAGATGACATGAGTAAACTGACATTATCCGAAGGCCATCCGGAAACGCCA
GTTGATGGGGACCTAGGGAAGCAAGATATCTGCTCATCTGAAGCCTCGTGGGGT
GATTTTGAATATGATGTAATGGGCCAGAATATCGATGAAGATTTCCTGAGAGAG
CCTGAACACTTCCTGTATGGTGGTGACCCTCCTTTGGAGGAAGATTCTCTGAAG
CAGTCGCTGGCACCGTACACACCTCCCTTTGATTTGTCTTATCTCACAGAACCT
GCCCAGAGTGCTGAAACAATAGAGGAAGCTGGGTCTCCAGAGGATGAATCTCTG
GGATGCAGAGCAGCAGAGATAGTGCTTTCTGCACTTCCTGATCGAAGAAGTGAG
GGAAACCAGGCTGAGACCAAAAACAGACTGCCTGGATCCCAGCTGGCTGTGCTG
CATATTCGTGAAGACCCTGAGTCCGTTTATTTGCCGGTAGGAGCAGGCTCCAAC
ATTTTGTCTCCATCAAACGTTGACTGGGAAGTAGAAACAGATAATTCTGATTTA
CCAGCAGGTGGAGACATAGGACCACCAAATGGTGCCAGCAAGGAAATACCAGAA
TTGGAAGAAGAAAAAACAATTCCTACCAAAGAGCCTGAGCAGATAAAATCAGAA
TACAAGGAAGAAAGATGTACAGAGAAGAATGAAGATCGTCATGCACTACACATG
GATTACATACTTGTAAACCGTGAAGAAAATTCACACTCAAAGCCAGAGACCTGT
GAAGAAAGAGAAAGCATAGCTGAATTAGAATTGTATGTAGGTTCCAAAGAAACA
GGGATGCAGGGAACTCAGTTAGCAAGCTTCCCAGACACATGTCAGCCAGCCTCC
TTAAATGAAAGAAAAGGTCTCTCTGCAGAGAAAATGTCTTCTAAAGGCGATACG
AGATCATCTTTTGAAAGCCCTGCACAAGACCAGAGTTGGATGTTCTTGGGCCAT
AGCGAGGTTGGTGATCCATCACTGGATGCCAGGGACTCAGGGCCTGGGTGGTCT
GGCAAGACTGTGGAGCCGTTCTCTGAACTCGGCTTGGGTGAGGGTCCCCAGCTG
CAGATTCTGGAAGAAATGAAGCCTCTAGAATCTTTAGCACTAGAGGAAGCCTCT
GGTCCAGTCAGCCAATCACAGAAGAGTAAGAGCCGAGGCAGGGCTGGCCCGGAT
GCAGTTACCCATGACAATGAATGGGAAATGCTTTCACCACAGCCTGTTCAGAAA
AACATGATCCCTGACACGGAAATGGAGGAGGAGACAGAGTTCCTTGAGCTCGGA
ACCAGGATATCAAGACCAAATGGACTACTGTCAGAGGATGTAGGAATGGACATC
CCCTTTGAAGAGGGCGTGCTGAGTCCCAGTGCTGCAGACATGAGGCCTGAACCT
CCTAATTCTCTGGATCTTAATGACACTCATCCTCGGAGAATCAAGCTCACAGCC
CCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTATTCTCTCTGATGAT
AACTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGAACTTGATACCCCC
GATGAAGCAGATTCTTTTGAGTACACTGGCCATGATCCCACAGCCAACAAAGAT
TCTGGCCAAGAGTCAGAGTCTATTCCAGAATATACGGCCGAAGAGGAACGGGAG
GACAACCGGCTTTGGAGGACAGTGGTCATTGGAGAACAAGAGCAGCGCATTGAC
ATGAAGGTCATCGAGCCCTACAGGAGAGTCATTTCTCACGGAGGAGATTCAGGA
TACTATGGGGACGGTCTAAATGCCATTATTGTGTTTGCCGCCTGTTTTCTGCCA
GACAGCAGTCGGGCGGATTACCACTATGTCATGGAAAATCTTTTCCTATATGTA
ATAAGTACTTTAGAGTTGATGGTAGCTGAAGACTATATGATTGTGTACTTGAAT
GGTGCAACCCCAAGAAGGAGGATGCCAGGGCTAGGCTGGATGAAGAAATGCTAC
CAGATGATTGACAGACGGTTGAGGAAGAATTTGAAATCATTCATCATTGTTCAT
CCATCTTGGTTCATCAGAACAATCCTTGCTGTGACACGACCTTTTATAAGTTCA

AAATTCAGCAGTAAAATTAAATATGTCAATAGCTTATCAGAACTCAGTGGGCTG
ATCCCAATGGATTGCATCCACATTCCAGAGAGCATCATCAATATTGACTTGAAG
CTGAAAGAAAAGCCT**TAG**TTGGCCATGCTGGAAGAAGAGGATGCTTTTCTGGTT
CATGGTTCTGTTGAAACATATCTACCTGAAAGAGACAGGGCTGATGTCACCTTT
TTCCACTTTGCACTACCTGGTGCCATTCTAAATTTCTAAGGGGAAAAATAGAAA
GTTTGTTTACTCTTAAGATATTTTATGAAATTGTGTGTACTTTCCTATTTTGCC
AATTATGTGCCTCAAAGATTTTAGTTGAGCCTTAGCAAGAAAGTAGGACCTTCC
ATTTCAATACTTCATTAACACGGTGTAGTGATACTTTGTCCCTTAGACTGGTGT
TTACCAGTAAGATACCTTTAATCCACTGTTAAGTATGAGTGGATTTGTTTCCAT
AGATTAGCTGGATTTCCTTTTGGTGATTGCATTAGGTTTAAAGTACACAGGTCT
CAACTCTCCCCAGGAAAGTTTCCCCTGTTTGACTCCACCTTTAAAATCCTAAGC
CTGACTAGGACAGCCACAAACCACACAAGGTGTAAAACCATCATCAGCTAAGTG
CCCGTTTTGTTCTTGTTTACCAGAATCTCCTTTAACTTCTCAAAGGGAAGCCGG
GCTTTCTAATCCACGTCAACTTTATTTTAGTTGTCAAATTGGGCATTATATTTT
ATGTAAATTGGTCTTTTAACATCATTTTCCTGATGAATGTTGGTGACCACCACA
TTGTGAAATTTAAGAATCCGTGTTGCATGTTTGGTAGCTCTCTGAGTTTCAGGC
CATAAACTCAGCTCCAGAGGTTACCTTTTAAGTGCCAAGAACTCAAGTGCAAGG
TGGCCTACTCAAAAATCATTTGGTAGCATTCAGTTATTCATGAATTCCTCTCTC
GCATGCATTATAAAAAGTGATCTGCTTTAAAACACCGTAATCTGATCATAGGCT
TAAAATTAAATATGAGTATTACTTTCATGTACAAAATATTTCCTTTATAGTCTT
CATATGCCCTTTAAAATGCCAACAAGATTTCAAGTCTGTAGGCCTCTAGTGAGG
TGGGGTGGCAAACCACAGCTAAGTCTCGCTCACCACTGCAAGCTAAGAATGGTT
TTTACATTTTGGGTTGGAAAAATTTTTTTTGAATATTTCATGACACATGAAAAT
TATTCAAATGTTAGTGCCGATAAATAAAGTGGTACTGAAACACAGCCACACAAA
CTTGTTTTTGTACTGTCTACAGCTACTTTCACACTACAGCCGCAGAGCTGAGCA
GTTCAGCAGACCGTATGTCCCACAATGCCTAAAACATTGACTATGTTTACAGAA
AAAGTTTGCTGACCCCTGCTCTAGCAAACGCATCCTTTCCTACTCCACCCCAAT
TTGTATTTAGATAGTTTCTCTAACAGAACGGACAAATGAGGCTGCAAACTAATT
TATTTTTGTCAAAAATCAATGTTTTGACATCCACAGACAGTGAAATAAAAGAAA
TGGCTTGCTGAAAAACATGAGGAGTCCTAGCCACAAAATCACTGCTTAGGTTGC
AATTGCCAAAATGAAGCCTTCTTAGAAGCACTTCTTTAGTATATACAGGTGTTG
GCTGAAGTCCGTGCCTCACTCTGGGAACCATTCTTAGTCTCCAGTGTCTCCTAT
TACAAAGAAGCTGGCAGAAATAAAAATGAAGGGGTGAGAGCGGTTCCACCCTAG
TCTCATGGTGGAAAATTCATTGGGGAGAGCTGTCCAGGATATTTGGAGTCCTGG
GTAGAAGGAGCTTGTAACTACTTTAAAGTCGACATCTTTGCACAGGTGATTGAG
TTTCTCTGACCTCATTGCTTCACCTCTGTCTCCTCCCGTCCTTCCGCACGTGCC
CACACACACGCAGTTCAGCCCTCTTTCCTCCATAAGCCTCCATCGTTTTCTCTT
TTCTCCTCTTGATCCTTTCAAGCGAGTATCTTGTTGAATTGTATGTTCTGTTGG
ATCTCCTCCTTCATAACATCTGGCTTGTTGGACAGAAAAACCCTACAGCCCACC
CCCTCCCACAGCCCACCTCCACTTTTGAAAGCCCAAATTACACCTCTCCCAGAA
CACAGTGTTGACGTAAATACAGTTACCCAATATTCCTGTTTGTTCACCTATTTG
CTACTTTCACTCAGTAGCATCCCATTTTGTAAAATGAATTCCATGGTCACCCTG
TCACAGGAAGTAATGAAAAATCCAGTGTTCAGTGTAGTGGTGCAAACCTGAGGG
CATAGAGCTGTTCATAGAGGGCTCTTGTTATAGCCAAACAGACACAGCAACAAT
CTCACCATTTATATATATATTTTTAACTTGTCCAGCTCATCTATGGAAAACTAC
TCAGGTGGTATGCTGTTTGAAGCCTCATCTTCCTACATGAAAATTATGGGCATT
TGTCCCAATGATTTTGTTTCAGCTGTTCTGTAGGCTGCATAACCACTCTGATAT
TTAGGTATCTGCTATTTTATTATCTTAAAAGACAAATTAATTTAATTGCATGTG

CTAGGGAAAAGCTACCATGTACATTCACCCCAAGTAAATAGAATCCTAGATGAA
TCCTAGAAAAATAATCCCTAAGCAGATAGGTAGACAGAGGTAAACATTCACATG
ATTTAGCTCTCTAGCTCTTGCACTCTGAACATTCTTGCTTTGGTTCTGACTTCT
GGGAACTGCTTTGCATTTCTCCTATAGATCTGTAGTTAAGGGAACCAAGGGGTC
ATTGGGGCAAAAGCATTGTTTCTCAAAGCTCCTTGATTAAGAGAAAGAACAGAA
ATTTGCACAGAAGATAGTGTCAAGGAGTGAGAAAGTTTGTTTGAGGGCAGTAGC
TCAGTGTGGAAGAAAATCCTGAAGTTTCTGTTGAAGCCATACAATGTTCTATGG
GGTTACTCTCTAAGACATTCTCTGAGGTGTGTGAGGAAGTCACTACTCCTAGCC
TTTGTTAAGATGTAATTTTAAATATTCAGTTATGGTACTATGTTTGCAACTCTC
GTCTTATCACAATGCCTCAGTAGTTTGTTCCCTTAGAAACATTTAGATGTGCAC
AAATTAATCTTTTATATATCTAAAGGTTTTTCTATCATGCATTGGATTGCTCAG
AATAAAGTGTCTGTTAGACTTCGTTTTGG

Sequence Translation: 3,057 amino acids

MEEFLQRAKSKLNRSKRLEKVHVVIGPKSCDLDSLISTFTYAYFLDKVSPPGVL
CLPVLNIPRTEFNYFTETRFILEELNISESFHIFRDGINLHQLNDEGKLSITLV
GSSVLASEDKTLESAVVKVINPVEQSDANVEFRESSSSLVLKEILQEAPELITE
QLAHRLRGSILFKWMTMESEKISEKQEEILSILEEKFPNLPPREDIINVLQETQ
FSAQGLSIEQTMLKDLKELSDGEIKVAISTVNMNLENCLFHSNITSDLKAFTDK
FGFDVLILFSSYLSEEQQPRRQIAVYSENMELCSQICCELEECQNPCLELEPFD
CGCDEILVYQQEDPSVTCDQVVLVVKEVINRRCPEMVSNSRTSSTEAVAGSAPL
SQGSSGIMELYGSDIEPQPSSVNFIENPPDLNDSNQAQVDANVDLVSPDSGLAT
IRSSRSSKESSVFLSDDSPVGEGAGPHHTLLPGLDSYSPIPEGAVAEEHAWSGE
HGEHFDLFNFDPAPMASGQSQQSSHSADYSPADDFFPNSDLSEGQLPAGPEGLD
GMGTNMSNYSSSSLLSGAGKDSLVEHDEEFVQRQDSPRDNSERNLSLTDFVGDE
SPSPERLKNTGKRIPPTPMNSLVESSPSTEEPASLYTEDMTQKATDTGHMGPPQ
THARCSSWWGGLEIDSKNIADAWSSSEQESVFQSPESWKEHKPSSIDRRASDSV
FQPKSLEFTKSGPWESEFGQPELGSNDIQDKNEESLPFQNLPIEKSPLPNTSPQ
GTNHLIEDFASLWHSGRSPTAMPEPWGNPTDDGEPAAVAPFPAWSAFGKEDHDE
ALKNTWNLHPTSSKTPSVRDPNEWAMAKSGFAFSSSELLDNSPSEINNEAAPEI
WGKKNNDSRDHIFAPGNPSSDLDHTWTNSKPPKEDQNGLVDPKTRGKVYEKVDS
WNLFEENMKKGGSDVLVPWEDSFLSYKCSDYSASNLGEDSVPSPLDTNYSTSDS
YTSPTFAGDEKETEHKPFAKEEGFESKDGNSTAEETDIPPQSLQQSSRNRISSG
PGNLDMWASPHTDNSSEINTTHNLDENELKTEHTDGKNISMEDDVGESSQSSYD
DPSMMQLYNETNRQLTLLHSSTNSRQTAPDSLDLWNRVILEDTQSTATISDMDN
DLDWDDCSGGAAIPSDGQTEGYMAEGSEPETRFTVRQLEPWGLEYQEANQVDWE
LPASDEHTKDSAPSEHHTLNEKSGQLIANSIWDSVMRDKDMSSFMLPGSSHITD
SEQRELPPEIPSHSANVKDTHFPDAPAASGTSESEALISHLDKQDTERETLQSD
AASLATRLENPGYFPHPDPWKGHGDGQSESEKEAQGATDRGHLDEEEVIASGVE
NASGISEKGQSDQELSSLVASEHQEICIKSGKISSLAVTFSPQTEEPEEVLEYE
EGSYNLDSRDVQTGMSADNLQPKDTHEKHLMSQRNSGETTETSDGMNFTKYVSV
PEKDLEKTEECNFLEPENVGGGPPHRVPRSLDFGDVPIDSDVHVSSTRSEITKN
LDVKGSENSLPGAGSSGNFDRDTISSEYTHSSASSPELNDSSVALSSWGQQPSS
GYQEENQGNWSEQNHQESELITTDGQVIVTKVKDLEKNRINEFEKSFDRKTPTF
LEIWNDSVDGDSFSSLSSPETGKYSEHSGTHQESNLIASYQEKNEHDISATVQP
EDARVISTSSGSDDDSVGGEESIEEEIQVANCHVAEDESRAWDSLNESNKFLVT

ADPKSENIYDYLDSSEPAENENKSNPFCDNQQSSPDPWTFSPLTETEMQITAVE
KEKRSSPETGTTGDVAWQISPKASFPKNEDNSQLEMLGFSADSTEWWKASPQEG
RLIESPFERELSDSSGVLEINSSVHQNASPWGVPVQGDIEPVETHYTNPFSDNH
QSPFLEGNGKNSHEQLWNIQPRQPDPDADKFSQLVKLDQIKEKDSREQTFVSAA
GDELTPETPTQEQCQDTMLPVCDHPDAAFTHAEENSCVTSNVSTNEGQETNQWE
QEKSYLGEMTNSSIATENFPAVSSPTQLIMKPGSEWDGSTPSEDSRGTFVPDIL
HGNFQEGGQLASAAPDLWIDAKKPFSLKADGENPDILTHCEHDSNSQASDSPDI
CHDSEAKEETEKHLSACMGPEVESSELCLTEPEIDEEPIYEPGREFVPSNAELD
SENATVLPPIGYQADIKGSSQPTSHKGSPEPSEINGDNSTGLQVSEKGASPDMA
PILEPVDRRIPRIENVATSIFVTHQEPTPEGDGSWISDSFSPESQPGARALFDG
DPHLSTENPALVPDALLASDTCLDISEAAFDHSFSDASGLNTSTGTIDDMSKLT
LSEGHPETPVDGDLGKQDICSSEASWGDFEYDVMGQNIDEDFLREPEHFLYGGD
PPLEEDSLKQSLAPYTPPFDLSYLTEPAQSAETIEEAGSPEDESLGCRAAEIVL
SALPDRRSEGNQAETKNRLPGSQLAVLHIREDPESVYLPVGAGSNILSPSNVDW
EVETDNSDLPAGGDIGPPNGASKEIPELEEEKTIPTKEPEQIKSEYKEERCTEK
NEDRHALHMDYILVNREENSHSKPETCEERESIAELELYVGSKETGMQGTQLAS
FPDTCQPASLNERKGLSAEKMSSKGDTRSSFESPAQDQSWMFLGHSEVGDPSLD
ARDSGPGWSGKTVEPFSELGLGEGPQLQILEEMKPLESLALEEASGPVSQSQKS
KSRGRAGPDAVTHDNEWEMLSPQPVQKNMIPDTEMEEETEFLELGTRISRPNGL
LSEDVGMDIPFEEGVLSPSAADMRPEPPNSLDLNDTHPRRIKLTAPNINLSLDQ
SEGSILSDDNLDSPDEIDINVDELDTPDEADSFEYTGHDPTANKDSGQESESIP
EYTAEEEREDNRLWRTVVIGEQEQRIDMKVIEPYRRVISHGGDSGYYGDGLNAI
IVFAACFLPDSSRADYHYVMENLFLYVISTLELMVAEDYMIVYLNGATPRRRMP
GLGWMKKCYQMIDRRLRKNLKSFIIVHPSWFIRTILAVTRPFISSKFSSKIKYV
NSLSELSGLIPMDCIHIPESIINIDLKLKEKP

B.2 PC-TSGC2 (GenBank accession number: FJ808773)

Total bp: 12,217

5’ UTR: 409 bp (1 – 409)

Kozak sequence: 407 – 413

ATG: 410 – 412

Coding sequence (in blue): 8202 bp (410- 8611)

Stop codon: 8609-8611

3’ UTR: 3606 bp (8612 – 12217)

Full-length cDNA sequence:

ACCCCGCTCGTCTTCCTTGGGCCGAGATTTTCCACTGCGCCCCTCCGAGTACCC
GGGTTCCAAACCCCTAGCCACGACATGGAAGAATTTTTGCAACGCGCCAAATCT
AAACTGAATCGAAGCAAACGCTTGGAGAAGGTCCATGTGGTTATTGGGCCTAAA
TCGTGTGACTTGGATTCTCTCATTTCTACCTTCACATATGCTTACTTTCTAGAC
AAGTGAAGACAAAACTTTAGAATCAGCAGTTGTCAAAGTCATTAATCCGGTTGA
GCAGAGCGATGCCAACGTTGAGTTCCGAGAGTCTTCCTCTTCTCTCGTGCTAAA
GGAGATTCTCCAGGAGGCTCCTGAGCTCATCACCGAGCAACTGGCTCATCGCCT
CAGAGGTAGCATTCTTTTCAAGTGGATGACC**ATG**GAATCAGAGAAGATCTCAGA
GAAGCAGGAGGAAATTCTTTCTATCCTGGAAGAAAAATTTCCTAACTTGCCTCC
AAGAGAGGACATCATCAACGTCCTACAGGAGACCCAGTTCAGTGCTCAGGGTTT
AAGTATTGAACAGACAATGTTGAAAGATCTAAAGGAGCTGTCAGATGGAGAAAT
AAAAGTGGCCATTAGTACTGTGAGCATGAACCTTGAGAATTGTCTATTTCACAG
CAATATTACCAGTGACTTGAAAGCATTTACAGACAAGTTTGGTTTTGATGTCCT
CATCCTGTTCTCCAGCTATCTGTCAGAGGAGCAGCAGCCGAGACGACAGATTGC
TGTGTACTCAGAAAACATGGAGCTGTGCAGTCAGATTTGCTGTGAGCTGGAAGA
GTGTCAGAACCCTTGCCTAGAACTGGAGCCCTTTGACTGTGGCTGTGATGAGAT
CCTGGTGTACCAACAAGAGGACCCTTCAGTGACTTGTGATCAGGTGGTTCTCGT
TGTCAAGGAAGTCATCAACAGGAGGTGTCCAGAGATGGTCTCCAATAGCCGGAC
ATCCTCAACAGAAGCCGTGGCAGGCAGTGCCCCCCTCTCCCAGGGGTCTTCTGG
GATTATGGAATTGTATGGTTCTGACATAGAGCCACAACCCAGCTCTGTGAATTT
CATAGAGAACCCTCCAGATCTCAATGATTCTAACCAGGCTCAGGTGGATGCCAA
TGTAGACCTTGTTAGCCCAGACAGCGGACTGGCTACCATTAGGAGCAGCCGCTC
ATCCAAGGAGAGCTCTGTTTTCCTCAGTGACGACAGCCCCGTGAGAGAAGGTGC

TGGGCCTCACCACACCCTTCTCCCAGGGCTTGACTCCTACAGCCCCATCCCTGA
AGGGGCGGTAGCGGAGGAACATGCATGGTCTGGAGAACACGGTGAGCACTTCGA
CCTCTTCAATTTTGACCCAGCACCCATGGCTTCTGGGCAGTCCCAGCAATCTTC
TCATTCTGCAGACTACTCCCCAGCAGATGACTTCTTCCCCAACAGCGACCTGTC
AGAAGGACAGCTCCCCGCTGGGCCTGAAGGACTTGATGGCATGGGAACCAACAT
GTCTAATTATTCATCCAGTTCACTTTTGTCAGGGGCTGGCAAAGATAGCCTTGT
GGAACATGATGAGGAGTTTGTCCAGAGACAAGACAGTCCCAGAGATAACTCTGA
AAGAAATTTGAGCCTGACAGATTTTGTGGGAGATGAATCCCCTTCCCCAGAAAG
GCTAAAAAATACTGGAAAGAGGATCCCACCAACACCCATGAATAGTTTAGTAGA
AAGCTCGCCATCCACTGAAGAACCAGCCTCACTCTATACCGAAGATATGACCCA
AAAAGCAACTGACACAGGTCACATGGGGCCACCTCAGACCCATGCACGGTGCAG
CAGCTGGTGGGGTGGTTTGGAAATTGACTCCAAAAATATTGCAGATGCGTGGAG
TTCCAGTGAACAGGAATCTGTTTTCCAGAGCCCTGAATCATGGAAAGAGCATAA
GCCAAGCTCCATTGATAGGAGAGCCTCAGATTCTGTATTTCAACCAAAGAGCCT
CGAATTTACAAAGTCAGGTCCCTGGGAGTCTGAATTTGGTCAGCCTGAACTGGG
TAGCAATGATATTCAAGACAAAAATGAGGAAAGCTTGCCGTTCCAGAACCTGCC
CATAGAGAAGTCACCTTTGCCAAATACATCTCCCCAAGGAACAAACCACCTGAT
AGAAGACTTTGCTTCTTTGTGGCATTCTGGTCGCTCTCCCACAGCCATGCCCGA
GCCCTGGGGAAATCCTACAGATGATGGTGAACCAGCAGCTGTGGCGCCATTCCC
AGCCTGGAGTGCATTTGGTAAAGAAGATCATGATGAAGCTTTAAAAAATACCTG
GAATTTGCACCCAACAAGCAGCAAGACACCTTCTGTTAGGGACCCGAATGAGTG
GGCCATGGCAAAAAGTGGGTTTGCCTTTTCTTCTTCAGAACTACTGGACAATTC
ACCCAGTGAGATAAACAATGAAGCAGCTCCAGAAATCTGGGGCAAGAAAAACAA
TGACTCCAGGGATCACATCTTTGCACCTGGAAATCCCAGTTCTGATCTGGATCA
CACATGGACTAATTCTAAGCCACCAAAAGAAGATCAGAATGGTTTAGTGGATCC
TAAAACTAGGGGCAAGGTATATGAAAAGGTAGACTCCTGGAACCTTTTTGAGGA
GAATATGAAGAAAGGAGGGTCAGATGTCCTAGTTCCTTGGGAAGATTCCTTCTT
ATCTTACAAATGTTCTGATTACAGTGCATCCAACCTAGGAGAAGATTCGGTGCC
TTCCCCCTTAGATACCAATTATTCCACCTCAGACTCTTACACATCACCAACATT
TGCTGGAGACGAAAAGGAAACTGAACACAAGCCATTTGCTAAAGAGGAAGGTTT
TGGGTCAAAAGATGGTAACTCCACGGCAGAGGAGACTGACATTCCTCCTCAGTC
ACTGCAACAGTCATCTCGAAATCGAATCAGTTCAGGTCCTGGGAACCTAGACAT
GTGGGCTTCACCTCATACAGATAACAGTTCTGAAATAAATACCACTCACAACCT
GGATGAAAATGAACTCAAGACAGAGCACACAGATGGTAAGAATATCTCCATGGA
GGATGACGTCGGGGAAAGCAGCCAGTCCAGTTACGACGACCCCAGCATGATGCA
ACTGTACAATGAAACAAACCGACAACTCACACTTTTGCACAGCAGCACCAACTC
CCGGCAGACGGCCCCTGACAGTCTCGACTTGTGGAACAGAGTGATTTTGGAGGA
TACTCAGTCCACTGCAACGATCTCAGATATGGACAATGATTTGGACTGGGATGA
CTGCAGTGGGGGTGCGGCAATCCCCAGTGATGGTCAAACAGAAGGATACATGGC
TGAAGGTTCCGAGCCGGAAACCCGATTTACAGTGAGACAGCTGGAACCCTGGGG
CTTGGAGTATCAGGAAGCAAATCAGGTAGATTGGGAGCTCCCTGCCTCTGATGA
GCATACCAAGGACAGTGCTCCCAGTGAACATCACACATTGAATGAGAAAAGTGG
GCAGCTAATTGCAAACAGTATTTGGGATTCTGTCATGAGAGATAAAGACATGTC
ATCATTCATGTTACCAGGCTCCTCACATATCACAGATTCAGAGCAAAGGGAATT
GCCTCCTGAAATCCCCAGCCATTCAGCAAATGTTAAAGACACTCATTCCCCAGA
TGCGCCAGCAGCCTCTGGAACCAGTGAATCAGAGGCACTTATATCTCATCTTGA
CAAGCAGGACACAGAGAGGGAAACCCTGCAAAGTGATGCAGCATCCTTGGCGAC
TAGGCTTGAGAATCCAGGGTATTTTCCACACCCAGATCCATGGAAAGGTCATGG

CGATGGACAAAGTGAAAGTGAGAAGGAAGCCCAGGGAGCCACTGACAGGGGGCA
CCTTGATGAAGAGGAGGTGATCGCCTCTGGTGTGGAGAATGCCTCAGGGATTTC
TGAAAAAGGGCAGAGTGACCAGGAACTGTCTTCTCTGGTTGCATCTGAACATCA
GGAAATCTGCATTAAATCAGGCAAAATCAGCTCTCTTGCTGTCACTTTCAGTCC
TCAAACCGAGGAACCAGAGGAAGTTTTAGAGTATGAGGAGGGGTCTTACAATCT
AGACTCCCGTGATGTGCAAACAGGGATGTCCGCAGATAACCTGCAGCCAAAAGA
TACCCATGAAAAACACCTCATGAGTCAAAGAAATTCAGGTGAAACTACTGAGAC
TTCAGATGGGATGAATTTCACAAAATATGTATCTGTACCTGAAAAGGATCTTGA
GAAAACTGAAGAATGTAACTTTCTAGAGCCAGAGAACGTGGGTGGAGGGCCACC
TCACAGAGTTCCCCGAAGTCTTGACTTTGGGGACGTCCCTATAGACAGTGATGT
GCATGTCAGCAGCACACGTTCTGAGATAACCAAAAATCTTGACGTTAAGGGGTC
TGAAAATAGCCTTCCAGGAGCCGGTTCGTCTGGAAATTTTGACAGAGATACTAT
TTCTAGTGAGTATACTCATTCAAGTGCATCAAGTCCTGAGTTAAATGACTCTTC
AGTTGCACTGTCCTCCTGGGGCCAGCAACCCAGTTCTGGGTATCAAGAAGAAAA
CCAAGGCAACTGGAGTGAACAGAATCACCAAGAATCTGAACTAATTACCACTGA
TGGCCAAGTAGAAATAGTTACCAAAGTGAAGGATTTAGAGAAAAACAGAATAAA
TGAGTTTGAAAAGAGCTTTGATCGCAAAACTCCTACATTTTTAGAGATCTGGAA
TGACTCAGTTGATGGTGATTCCTTTTCCTCTTTATCCAGTCCTGAAACAGGCAA
ATATTCTGAACATTCAGGGACACATCAGGAAAGCAATCTAATTGCTAGCTACCA
GGAGAAAAATGAACATGACATTTCTGCAACTGTGCAGCCAGAGGATGCCAGGGT
CATTTCAACAAGCTCAGGTTCTGATGATGACAGTGTCGGTGGTGAAGAGTCAAT
AGAGGAAGAGATCCAGGTGGCCAACTGCCACGTTGCTGAGGATGAATCCAGAGC
TTGGGATTCATTGAATGAATCTAATAAGTTCTTGGTCACAGCTGATCCTAAGTC
TGAAAATATTTATGACTACCTAGACAGCTCAGAGCCAGCAGAGAATGAGAATAA
GTCAAACCCATTCTGTGACAATCAACAAAGCAGCCCTGATCCCTGGACTTTCTC
ACCATTAACGGAGACTGAAATGCAGATTACAGCAGTGGAGAAGGAGAAGAGATC
TTCTCCAGAAACAGGGACAACAGGAGATGTTGCATGGCAAATATCTCCCAAAGC
TTCGTTCCCAAAGAACGAAGATAATTCTCAACTGGAAATGCTGGGCTTCTCAGC
TGATAGCACTGAGTGGTGGAAGGCCTCACCCCAGGAAGGGAGACTAATTGAAAG
TCCATTTGAAAGGGAGCTGTCTGACTCCAGTGGTGTGTTGGAGATAAATTCTTC
AGTACACCAAAATGCCAGTCCCTGGGGAGTACCAGTTCAGGGTGATATTGAGCC
CGTGGAAACACACTATACTAATCCTTTTAGTGACAACCATCAGTCACCCTTTCT
GGAAGGTAATGGGAAGAACTCCCATGAGCAACTCTGGAACATTCAACCAAGGCA
GCCAGACCCAGATGCTGACAAGTTCAGCCAGCTTGTAAAATTAGACCAAATTAA
AGAAAAGGACTCAAGAGAGCAAACCTTTGTGTCTGCTGCTGGTGATGAGCTGAC
TCCTGAAACACCTACCCAAGAGCAGTGTCAGGACACCATGCTGCCAGTCTGTGA
TCATCCGGACGCAGCCTTTACTCACGCAGAGGAAAATAGTTGTGTTACATCTAA
TGTTTCAACTAATGAAGGTCAAGAAACAAATCAGTGGGAACAAGAAAAATCATA
CCTAGGTGAGATGACAAATTCAAGCATTGCCACAGAAAATTTTCCTGCTGTCAG
TTCTCCCACCCAACTGATAATGAAGCCAGGCTCTGAATGGGATGGCTCTACCCC
AAGTGAGGACTCCCGAGGTACCTTTGTGCCAGATATTTTACATGGCAACTTTCA
AGAGGGTGGGCAGCTGGCCTCTGCCGCGCCTGACTTGTGGATAGATGCTAAGAA
GCCCTTCAGTTTGAAAGCAGATGGTGAGAATCCTGATATCCTGACGCACTGCGA
ACATGACAGCAATTCTCAGGCTTCCGACAGCCCTGATATATGTCACGATTCTGA
AGCAAAGGAAGAGACTGAGAAGCACCTCAGTGCTTGCATGGGACCTGAAGTGGA
ATCCAGTGAGCTTTGTCTCACTGAGCCAGAGATAGATGAAGAACCCATTTATGA
GCCTGGACGGGAGTTTGTCCCATCCAATGCAGAACTCGATTCTGAAAACGCAAC
TGTGCTGCCTCCAATTGGCTATCAAGCAGACATAAAGGGCTCCTCTCAGCCCAC

CTCTCATAAAGGTTCACCTGAACCTTCTGAAATAAACGGTGACAACAGTACAGG
TTTACAAGTATCAGAAAAAGGAGCCAGCCCAGATATGGCACCAATTTTGGAACC
AGTTGACAGAAGAATCCCAAGGATTGAAAATGTGGCAACTAGCATTTTTGTAAC
TCACCAAGAGCCAACTCCAGAAGGTGACGGTTCTTGGATATCAGACAGCTTTTC
TCCTGAAAGTCAGCCTGGTGCAAGAGCTTTGTTTGATGGTGATCCACATTTATC
CACAGAGAATCCTGCCTTGGTTCCTGATGCTTTGCTAGCCTCAGACACTTGTCT
GGATATAAGCGAAGCTGCCTTTGACCACAGTTTCAGCGATGCCTCAGGTCTCAA
CACATCCACGGGAACAATAGATGACATGAGTAAACTGACATTATCCGAAGGCCA
TCCGGAAACGCCAGTTGATGGGGACCTAGGGAAGCAAGATATCTGCTCATCTGA
AGCCTCGTGGGGTGATTTTGAATATGATGTAATGGGCCAGAATATCGATGAAGA
TTTCCTGAGAGAGCCTGAACACTTCCTGTATGGTGGTGACCCTCCTTTGGAGGA
AGATTCTCTGAAGCAGTCGCTGGCACCGTACACACCTCCCTTTGATTTGTCTTA
TCTCACAGAACCTGCCCAGAGTGCTGAAACAATAGAGGAAGCTGGGTCTCCAGA
GGATGAATCTCTGGGATGCAGAGCAGCAGAGATAGTGCTTTCTGCACTTCCTGA
TCGAAGAAGTGAGGGAAACCAGGCTGAGACCAAAAACAGACTGCCTGGATCCCA
GCTGGCTGTGCTGCATATTCGTGAAGACCCTGAGTCCGTTTATTTGCCGGTAGG
AGCAGGCTCCAACATTTTGTCTCCATCAAACGTTGACTGGGAAGTAGAAACAGA
TAATTCTGATTTACCAGCAGGTGGAGACATAGGACCACCAAATGGTGCCAGCAA
GGAAATACCAGAATTGGAAGAAGAAAAAACAATTCCTACCAAAGAGCCTGAGCA
GATAAAATCAGAATACAAGGAAGAAAGATGTACAGAGAAGAATGAAGATCGTCA
TGCACTACACATGGATTACATACTTGTAAACCGTGAAGAAAATTCACACTCAAA
GCCAGAGACCTGTGAAGAAAGAGAAAGCATAGCTGAATTAGAATTGTATGTAGG
CTCCAAAGAAACAGGGCTGCAGGGAACTCAGTTAGCAAGCTTCCCAGACACATG
TCAGCCAGCCTCCTTAAATGAAAGAAAAGGTCTCTCTGCAGAGAAAATGTCTTC
TAAAGGCGATACGAGATCATCTTTTGAAAGCCCTGCACAAGACCAGAGTTGGAT
GTTCTTGGGCCATAGTGAGGTTGGTGATCCATCACTGGATGCCAGGGACTCAGG
GCCTGGGTGGTCTGGCAAGACTGTGGAGCCGTTCTCTGAACTCGGCTTGGGTGA
GGGTCCCCAGCTGCAGATTCTGGAAGAAATGAAGCCTCTAGAATCTTTAGCACT
AGAGGAAGCCTCTGGTCCAGTCAGCCAATCACAGAAGAGTAAGAGCCGAGGCAG
GGCTGGCCCGGGTGCAGTTACCCATGACAATGAATGGGAAATGCTTTCACCACA
GCCTGTTCAGAAAAACATGATCCCTGACACGGAAATGGAGGAGGAGACAGAGTT
CCTTGAGCTCGGAACCAGGATATCAAGACCAAATGGACTACTGTCAGAGGATGT
AGGAATGGACATCCCCTTTGAAGAGGGCGTGCTGAGTCCCAGTGCTGCAGACAT
GAGGCCTGAACCTCCTAATTCTCTGGATCTTAATGACACTCATCCTCGGAGAAT
CAAGCTCACAGCCCCAAATATCAATCTTTCTCTGGACCAAAGTGAAGGATCTAT
TCTCTCTGATGATAACTTGGACAGTCCAGATGAAATTGACATCAATGTGGATGA
ACTTGATACCCCCGATGAAGCAGATTCTTTTGAGTACACTGGCCATGATCCCAC
AGCCAACAAAGATTCTGGCCAAGAGTCAGAGTCTATTCCAGAATATACGGCCGA
AGAGGAACGGGAGGACAACCGGCTTCGGAGGACAGTGGTCATTGGAGAACAAGA
GCAGCGCATTGACATGAAGGTCATCGAGCCCTACAGGAGAGTCATTTCTCACGG
AGGAGATTCAGGTAACCTTCTC**TGA**AGTCTTGCTGGACAAATGACCCTGGAAAG
AATGAGGAGAAGAGTTGCTGGCTGGTGTGTGTGTGTGTATGTTTTGGTTGCGAC
TGTGACATCTTTCCTTCTGTCTCCATTCCATTGCTGGCCAGCAAACTTGTGTTA
CATCCCTTTCATGCCTGTCTGTGTCACTCAGATATGTAATAAGTACTTTAGAGT
TGATGGTAGCTGAAGACTATATGATTGTGTACTTGAATGGTGCAACCCCAAGAA
GGAGGATGCCAGGGCTAGGCTGGATGAAGAAATGCTACCAGATGATTGACAGAC
GGTTGAGGAAGAATTTGAAATCATTCATCATTGTTCATCCATCTTGGTTCATCA
GAACAATCCTTGCTGTGACACGACCTTTTATAAGTTCAAAATTCAGCAGTAAAA

TTAAATATGTCAATAGCTTATCAGAACTCAGTGGGCTGATCCCAATGGATTGCA
TCCACATTCCAGAGAGCATCATCAATATTGACTTGAAGCTGAAAGAAAAGCCTT
AGTTGGCCATGCTGGAAGAAGAGGATGCTTTTCTGGTTCATGGTTCTGTTGAAA
CATATCTACCTGAAAGAGACAGGGCTGATGTTACCTTTTTCCACTTTGCACTAC
CTGGTGCCATTCTAAATTTCTAAGGGGAAAAATAGAAAGTTTGTTTACTCTTAA
GATATTTTATGAAATTGTGTGTACTTTCCTATTTTGCCAATTATGTGCCTCAAA
GATTTTAGTTGAGCCTTAGCAAGAAAGTAGGACCTTCCATTTCAATACTTCATT
AACACGGTGTAGTGATACTTTGTCCCTTAGACTGGTGTTTACCAGTAAGATACC
TTTAATCCACTGTTAAGTATGAGTGGATTTGTTTCCATAGATTAGCTGGATTTC
CTTTTGGTGATTGCATTAGGTTTAAAGTACACAGGTCTCAACTCTCCCCAGGAA
AGTTTCCCCTGTTTGACTCCACCTTTAAAATCCTAAGCCTGACTAGGACAGCCA
CAAACCACACAAGGTGTAAAACCATCATCAGCTAAGTGCCCGTTTTGTTCTTGT
TTACCAGAATCTCCTTTAACTTCTCAAAGGGAAGCCGGGCTTTCTAATCCACGT
CAACTTTATTTTAGTTGTCAAATTGGGCATTATATTTTATGTAAATTGGTCTTT
TAACATCATTTTCCTGATGAATGTTGGTGACCACCACATTGTGAAATTTAAGAA
TCCGTGTTGCATGTTTGGTAGCTCTCTGAGTTTCAGGCCATAAACTCAGCTCCA
GAGGTTACCTTTTAAGTGCCAAGAACTCAAGTGCAAGGTGGCCTACTCAAAAAT
CATTTGGTAGCATTCAGTTATTCATGAATTCCTCTCTCGCATGCATTATAAAAA
GTGATCTGCTTTAAAACACCGTAATCTGATCATAGGCTTAAAATTAAATATGAG
TATTACTTTCATGTACAAAATATTTCCTTTATAGTCTTCATATGCCCTTTAAAA
TGCCAACAAGATTTCAAGTCTGTAGGCCTCTAGTGAGGTGGGGTGGCAAACCAC
AGCTAAGTCTCGCTCACCACTGCAAGCTAAGAATGGTTTTTACATTTTGGGTTG
GAAAAATTTTTTTTGAATATTTCATGACACATGAAAATTATTCAAATGTTAGTG
CCGATAAATAAAGTGGTACTGAAACACAGCCACACAAACTTGTTTTTGTACTGT
CTACAGCTACTTTCACACTACAGCCGCAGAGCTGAGCAGTTCAGCAGACCGTAT
GTCCCACAATGCCTAAAACATTGACTATGTTTACAGAAAAAGTTTGCTGACCCC
TGCTCTAGCAAACGCATCCTTTCCTACTCCACCCCAATTTGTATTTAGATAGTT
TCTCTAACAGAACGGACAAATGAGGCTGCAAACTAATTTATTTTTGTCAAAAAT
CAATGTTTTGACATCCACAGACAGTGAAATAAAAGAAATGGCTTGCTGAAAAAC
ATGAGGAGTCCTAGCCACAAAATCACTGCTTAGGTTGCAATTGCCAAAATGAAG
CCTTCTTAGAAGCACTTCTTTAGTATATACAGGTGTTGGCTGAAGTCCGTGCCT
CACTCTGGGAACCATTCTTAGTCTCCAGTGTCTCCTATTACAAAGAAGCTGGCA
GAAATAAAAATGAAGGGGTGAGAGCGGTTCCACCCTAGTCTCATGGTGGAAAAT
TCATTGGGGAGAGCTGTCCAGGATATTTGGAGTCCTGGGTAGAAGGAGCTTGTA
ACTACTTTAAAGTCGACATCTTTGCACAGGTGATTGAGTTTCTCTGACCTCATT
GCTTCACCTCTGTCTCCTCCCGTCCTTCCGCACGTGCCCACACACACGCAGTTC
AGCCCTCTTTCCTCCATAAGCCTCCATCGTTTTCTCTTTTCTCTTCTTGATCCT
TTCAAGCGAGTATCTTGTTGAATTGTATGTTCTGTTGGATCTCCTCCTTCATAA
CATCTGGCTTGTTGGACAGAAAAACCCTACAGCCCACCCCCTCCCACAGCCCAC
CTCCACTTTTGAAAGCCCAAATTACACCTCTCCCAGAACACAGTGTTGACGTAA
ATACAGTTACCCAATATTCCTGTTTGTTCACCTATTTGCTACTTTCACTCAGTA
GCATCCCATTTTGTAAAATGAATTCCATGGTCACCCTGTCACAGGAAGTAATGA
AAAATCCAGTGTTCAGTGTAGTGGTGCAAACCTGAGGGCATAGAGCTGTTCATA
GAGGGCTCTTGTTATAGCCAAACAGACACAGCAACAATCTCACCATTTATATAT
ATATTTTTAACTTGTCCAGCTCATCTATGGAAAACTACTCAGGTGGTATGCTGT
TTGAAGCCTCATCTTCCTACATGAAAATTATGGGCATTTGTCCCAATGATTTTG
TTTCAGCTGTTCTGTAGGCTGCATAACCACTCTGATATTTAGGTATCTGCTATT
TTATTATCTTAAAAGACAAATTAATTTAATTGCATGTGCTAGGGAAAAGCTACC

ATGTACATTCACCCCAAGTAAATAGAATCCTAGATGAATCCTAGAAAAATAATC
CCTAAGCAGATAGGTAGACAGAGGTAAACATTCACATGATTTAGCTCTCTAGCT
CTTGCACTCTGAACATTCTTGCTTTGGGTCTGACTTCTGGGAACTGCTTTGCAT
TTCTCCTATAGATCTGTAGTTAAGGGAACCAAGGGGTCATTGGGGCAAAAGCAT
TGTTTCTCAAAGCTCCTTGATTAAGAGAAAGAACAGAAATTTGCACAGAAGATA
GTGTCAAGGAGTGAGAAAGTTTGTTTGAGGGCAGTAGCTCAGTGTGGAAGAAAA
TCCTGAAGTTTCTGTTGAAGCCATACAATGTTCTATGGGGTTACTCTCTAAGAC
ATTCTCTGAGGTGTGTGAGGAAGTCACTACTCCTAGCCTTTGTTAAGATGTAAT
TTTAAATATTCAGTTATGGTACTATGTTTGCAACTCTCGTCTTATCACAATGCC
TCAGTAGTTTGTTCCCTTAGAAACATTTAGATGTGCACAAATTAATCTTTTATA
TATCTAAAGGTTTTTCTATCATGCATTGGATTGCTCAGAATAAAGTGTCTGTTA
GACTTCGTTTTGG

Sequence translation: 2,733 amino acids

MESEKISEKQEEILSILEEKFPNLPPREDIINVLQETQFSAQGLSIEQTMLKDL
KELSDGEIKVAISTVSMNLENCLFHSNITSDLKAFTDKFGFDVLILFSSYLSEE
QQPRRQIAVYSENMELCSQICCELEECQNPCLELEPFDCGCDEILVYQQEDPSV
TCDQVVLVVKEVINRRCPEMVSNSRTSSTEAVAGSAPLSQGSSGIMELYGSDIE
PQPSSVNFIENPPDLNDSNQAQVDANVDLVSPDSGLATIRSSRSSKESSVFLSD
DSPVREGAGPHHTLLPGLDSYSPIPEGAVAEEHAWSGEHGEHFDLFNFDPAPMA
SGQSQQSSHSADYSPADDFFPNSDLSEGQLPAGPEGLDGMGTNMSNYSSSSLLS
GAGKDSLVEHDEEFVQRQDSPRDNSERNLSLTDFVGDESPSPERLKNTGKRIPP
TPMNSLVESSPSTEEPASLYTEDMTQKATDTGHMGPPQTHARCSSWWGGLEIDS
KNIADAWSSSEQESVFQSPESWKEHKPSSIDRRASDSVFQPKSLEFTKSGPWES
EFGQPELGSNDIQDKNEESLPFQNLPIEKSPLPNTSPQGTNHLIEDFASLWHSG
RSPTAMPEPWGNPTDDGEPAAVAPFPAWSAFGKEDHDEALKNTWNLHPTSSKTP
SVRDPNEWAMAKSGFAFSSSELLDNSPSEINNEAAPEIWGKKNNDSRDHIFAPG
NPSSDLDHTWTNSKPPKEDQNGLVDPKTRGKVYEKVDSWNLFEENMKKGGSDVL
VPWEDSFLSYKCSDYSASNLGEDSVPSPLDTNYSTSDSYTSPTFAGDEKETEHK
PFAKEEGFGSKDGNSTAEETDIPPQSLQQSSRNRISSGPGNLDMWASPHTDNSS
EINTTHNLDENELKTEHTDGKNISMEDDVGESSQSSYDDPSMMQLYNETNRQLT
LLHSSTNSRQTAPDSLDLWNRVILEDTQSTATISDMDNDLDWDDCSGGAAIPSD
GQTEGYMAEGSEPETRFTVRQLEPWGLEYQEANQVDWELPASDEHTKDSAPSEH
HTLNEKSGQLIANSIWDSVMRDKDMSSFMLPGSSHITDSEQRELPPEIPSHSAN
VKDTHSPDAPAASGTSESEALISHLDKQDTERETLQSDAASLATRLENPGYFPH
PDPWKGHGDGQSESEKEAQGATDRGHLDEEEVIASGVENASGISEKGQSDQELS
SLVASEHQEICIKSGKISSLAVTFSPQTEEPEEVLEYEEGSYNLDSRDVQTGMS
ADNLQPKDTHEKHLMSQRNSGETTETSDGMNFTKYVSVPEKDLEKTEECNFLEP
ENVGGGPPHRVPRSLDFGDVPIDSDVHVSSTRSEITKNLDVKGSENSLPGAGSS
GNFDRDTISSEYTHSSASSPELNDSSVALSSWGQQPSSGYQEENQGNWSEQNHQ
ESELITTDGQVEIVTKVKDLEKNRINEFEKSFDRKTPTFLEIWNDSVDGDSFSS
LSSPETGKYSEHSGTHQESNLIASYQEKNEHDISATVQPEDARVISTSSGSDDD
SVGGEESIEEEIQVANCHVAEDESRAWDSLNESNKFLVTADPKSENIYDYLDSS
EPAENENKSNPFCDNQQSSPDPWTFSPLTETEMQITAVEKEKRSSPETGTTGDV
AWQISPKASFPKNEDNSQLEMLGFSADSTEWWKASPQEGRLIESPFERELSDSS

GVLEINSSVHQNASPWGVPVQGDIEPVETHYTNPFSDNHQSPFLEGNGKNSHEQ
LWNIQPRQPDPDADKFSQLVKLDQIKEKDSREQTFVSAAGDELTPETPTQEQCQ
DTMLPVCDHPDAAFTHAEENSCVTSNVSTNEGQETNQWEQEKSYLGEMTNSSIA
TENFPAVSSPTQLIMKPGSEWDGSTPSEDSRGTFVPDILHGNFQEGGQLASAAP
DLWIDAKKPFSLKADGENPDILTHCEHDSNSQASDSPDICHDSEAKEETEKHLS
ACMGPEVESSELCLTEPEIDEEPIYEPGREFVPSNAELDSENATVLPPIGYQAD
IKGSSQPTSHKGSPEPSEINGDNSTGLQVSEKGASPDMAPILEPVDRRIPRIEN
VATSIFVTHQEPTPEGDGSWISDSFSPESQPGARALFDGDPHLSTENPALVPDA
LLASDTCLDISEAAFDHSFSDASGLNTSTGTIDDMSKLTLSEGHPETPVDGDLG
KQDICSSEASWGDFEYDVMGQNIDEDFLREPEHFLYGGDPPLEEDSLKQSLAPY
TPPFDLSYLTEPAQSAETIEEAGSPEDESLGCRAAEIVLSALPDRRSEGNQAET
KNRLPGSQLAVLHIREDPESVYLPVGAGSNILSPSNVDWEVETDNSDLPAGGDI
GPPNGASKEIPELEEEKTIPTKEPEQIKSEYKEERCTEKNEDRHALHMDYILVN
REENSHSKPETCEERESIAELELYVGSKETGLQGTQLASFPDTCQPASLNERKG
LSAEKMSSKGDTRSSFESPAQDQSWMFLGHSEVGDPSLDARDSGPGWSGKTVEP
FSELGLGEGPQLQILEEMKPLESLALEEASGPVSQSQKSKSRGRAGPGAVTHDN
EWEMLSPQPVQKNMIPDTEMEEETEFLELGTRISRPNGLLSEDVGMDIPFEEGV
LSPSAADMRPEPPNSLDLNDTHPRRIKLTAPNINLSLDQSEGSILSDDNLDSPD
EIDINVDELDTPDEADSFEYTGHDPTANKDSGQESESIPEYTAEEEREDNRLRR
TVVIGEQEQRIDMKVIEPYRRVISHGGDSGNLL

BIBLIOGRAPHY

Amaral PP, Dinger ME, Mercer TR, Mattick JS. (2008). The eukaryotic genome as an RNA machine. *Science* 319: 1787-1789.

Aravind L, Koonin EV. (1998). A novel family of predicted phosphoesterases includes Drosophila prune protein and bacterial RecJ exonuclease. *Trends Biochem Sci* 23: 17-19.

Basile JR, Gavard J, Gutkind JS. (2007). Plexin-B1 utilizes RhoA and Rho kinase to promote the integrin-dependent activation of Akt and ERK and endothelial cell motility. *J Biol Chem* 282: 34888-34895.

Brantl S. (2007). Regulatory mechanisms employed by cis-encoded antisense RNAs. *Curr Opin Microbiol* 10:102-109

Bussemakers MJ, van Bokhoven A, Verhaegh GW, Smit FP, Karthaus HF, Schalken JA, Debruyne FM, Ru N, Isaacs WB. (1999). DD3: a new prostate-specific gene, highly overexpressed in prostate cancer. *Cancer Res* 59: 5975-5979.

Carninci P, Yasuda J, Hayashizaki Y. (2008). Multifaceted mammalian transcriptome. *Curr Opin Cell Biol* 20: 274-280.

Carthew RW, Sontheimer EJ. (2009). Origins and Mechanisms of miRNAs and siRNAs. *Cell* 136: 642-655.

Chu CY, Rana TM. (2007). Small RNAs: regulators and guardians of the genome. *Cell Physiol* 213:412-419.

Clarke RA, Zhao Z, Guo AY, Roper K, Teng L, Fang ZM, Samaratunga H, Lavin MF, Gardiner RA. (2009). New genomic structure for prostate

cancer specific gene PCA3 within BMCC1: implications for prostate cancer detection and progression. *PLoS ONE* 4: e4995.

Coleman ML, Marshall CJ, Olson MF. (2004). RAS and RHO GTPases in G1-phase cell-cycle regulation. *Nat Rev Mol Cell Biol* 5: 355-366.

D'Angelo A, Garzia L, André A, Carotenuto P, Aglio V, Guardiola O. (2004). Prune cAMP phosphodiesterase binds nm23-H1 and promotes cancer metastasis. *Cancer Cell* 5: 137-149.

de Kok JB, Verhaegh GW, Roelofs RW, Hessels D, Kiemeney LA, Aalders TW, Swinkels DW, Schalken JA. (2002). DD3(PCA3), a very sensitive and specific marker to detect prostate tumors. *Cancer Res* 62: 2695-2698.

de la Taille A. (2007). Progensa PCA3 test for prostate cancer detection. *Expert Rev Mol Diagn* 7: 491-497.

Del Re DP, Miyamoto S, Brown JH. (2008). Focal adhesion kinase as a RhoA-activable signaling scaffold mediating Akt activation and cardiomyocyte protection. *J Biol Chem* 283: 35622-35629.

Faghihi MA, Modarresi F, Khalil AM, Wood DE, Sahagan BG, Morgan TE, Finch CE, St Laurent G 3rd, Kenny PJ, Wahlestedt C. (2008). Expression of a noncoding RNA is elevated in Alzheimer's disease and drives rapid feed-forward regulation of beta-secretase. *Nat Med* 14: 723-730.

Faghihi MA, Wahlestedt C. (2009). Regulatory roles of natural antisense transcripts. *Nat Rev Mol Cell Biol* 10: 637-43.

Farazi TA, Juranek SA, Tuschl T. (2008). The growing catalog of small RNAs and their association with distinct Argonaute/Piwi family members. *Development* 135: 1201-1214.

Galasso A, Zollo M. (2009). The Nm23-H1-h-Prune complex in cellular physiology: a 'tip of the iceberg' protein network perspective. *Mol Cell Biochem* 329: 149-59.

Gleason DF, Mellinger GT. (1974). Prediction of prognosis for prostatic adenocarcinoma by combined histological grading and clinical staging. *J Urol* 111: 58-64

Groskopf J, Aubin SM, Deras IL, Blase A, Bodrug S, Clark C, Brentano S, Mathis J, Pham J, Meyer T, Cass M, Hodge P, Macairan ML, Marks LS, Rittenhouse H. (2006). APTIMA PCA3 molecular urine test: development of a method to aid in the diagnosis of prostate cancer. *Clin Chem* 52: 1089-1095.

Hartsough MT, Steeg PS. (2000). Nm23/nucleoside diphosphate kinase in human cancers. *J Bioenerg Biomembr* 32: 301-308.

Igawa M, Rukstalis DB, Tanabe T, Chodak GW. (1994). High levels of nm23 expression are related to cell proliferation in human prostate cancer. *Cancer Res* 54: 1313-1318.

Jensen SL, Wood DP Jr, Banks ER, Veron M, Lascu I, McRoberts JW, Rangnekar VM. (1996). Increased levels of nm23 H1/nucleoside diphosphate kinase A mRNA associated with adenocarcinoma of the prostate. *World J Urol* 14: S21-25.

Kawaji H, Hayashizaki Y. (2008). Exploration of small RNAs. *PLoS Genet* 4:e22.

Katayama S, Tomaru Y, Kasukawa T, Waki K, Nakanishi M, Nakamura M, Nishida H, Yap CC, Suzuki M, Kawai J, Suzuki H, Carninci P, Hayashizaki Y, Wells C, Frith M, Ravasi T, Pang KC, Hallinan J, Mattick J, Hume DA, Lipovich L, Batalov S, Engström PG, Mizuno Y, Faghihi MA, Sandelin A, Chalk AM, Mottagui-Tabar S, Liang Z, Lenhard B, Wahlestedt C. (2005). Antisense transcription in the mammalian transcriptome. *Science* 309: 1564-1566.

Keegan LP, Leroy A, Sproul D, O'Connell MA. (2004). Adenosine deaminases acting on RNA (ADARs): RNA-editing enzymes. *Genome Biol* 5:209

Kim YI, Park S, Jeoung DI, Lee H. (2003). Point mutations affecting the oligomeric structure of Nm23-H1 abrogates its inhibitory activity on colonization and invasion of prostate cancer cells. *Biochem Biophys Res Commun* 307: 281-289.

Kimura K, Wakamatsu A, Suzuki Y, Ota T, Nishikawa T, Yamashita R, Yamamoto J, Sekine M, Tsuritani K, Wakaguri H, Ishii S, Sugiyama T, Saito K, Isono Y, Irie R, Kushida N, Yoneyama T, Otsuka R, Kanda K, Yokoi T, Kondo H, Wagatsuma M, Murakawa K, Ishida S, Ishibashi T, Takahashi-Fujii A, Tanase T, Nagai K, Kikuchi H, Nakai K, Isogai T, Sugano S. (2006). Diversification of transcriptional modulation: large-

scale identification and characterization of putative alternative promoters of human genes. *Genome Res* 16: 55-65.

Konishi N, Nakaoka S, Tsuzuki T, Matsumoto K, Kitahori Y, Hiasa Y. (1993). Expression of nm23-H1 and nm23-H2 proteins in prostate carcinoma. *Jpn J Cancer Res* 84: 1050-4.

Larisch S, Yi Y, Lotan R, Kerner H, Eimerl S, Tony Parks W, Gottfried Y, Birkey Reffey S, de Caestecker MP, Danielpour D, Book-Melamed N, Timberg R, Duckett CS, Lechleider RJ, Steller H, Orly J, Kim SJ, Roberts AB. (2000). A novel mitochondrial septin-like protein, ARTS, mediates apoptosis dependent on its P-loop motif. *Nat Cell Biol* 2: 915-921.

Lee HY, Lee H. (1999). Inhibitory activity of nm23-H1 on invasion and colonization of human prostate carcinoma cells is not mediated by its NDP kinase activity. *Cancer Lett* 145: 93-99.

Lee RC, Feinbaum RL, Ambros V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. *Cell* 75: 843-854.

Lifschytz E, Falk R. (1969). The action of the gene prune (pn) in Drosophila melanogaster. *Genet Res* 14: 53-61.

Livak KJ, Schmittgen TD. (2001). Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. *Methods* 25: 402-408.

Louro R, Nakaya HI, Amaral PP, Festa F, Sogayar MC, da Silva AM, Verjovski-Almeida S, Reis EM. (2007). Androgen responsive intronic non-coding RNAs. *BMC Biol* 5: 4.

Louro R, Smirnova AS, Verjovski-Almeida S. (2009). Long intronic noncoding RNA transcription: expression noise or expression choice? *Genomics* 93: 291-298.

Low BC, Seow KT, Guy GR. (2000). The BNIP-2 and Cdc42GAP homology domain of BNIP-2 mediates its homophilic association and heterophilic interaction with Cdc42GAP. *J Biol Chem* 275: 37742-37751.

Machida T, Fujita T, Ooo ML, Ohira M, Isogai E, Mihara M, Hirato J, Tomotsune D, Hirata T, Fujimori M, Adachi W, Nakagawara A. (2006). Increased expression of proapoptotic BMCC1, a novel gene with the BNIP2 and Cdc42GAP homology (BCH) domain, is associated with favorable prognosis in human neuroblastomas. *Oncogene* 25: 1931-1942.

Makarov DV, Loeb S, Getzenberg RH, Partin AW. (2009). Biomarkers for prostate cancer. *Annu Rev Med* 60: 139-151.

Mendes Soares LM, Valcárcel J. (2006). The expanding transcriptome: the genome as the 'Book of Sand'. *EMBO J* 25: 923-931.

Mercer TR, Dinger ME, Mattick JS. (2009). Long non-coding RNAs: insights into functions. *Nat Rev Genet* 10: 155-159.

Middelhaufe S, Garzia L, Ohndorf UM, Kachholz B, Zollo M, Steegborn C. (2007). Domain mapping on the human metastasis regulator protein h-Prune reveals a C-terminal dimerization domain. *Biochem J* 407: 199-205.

Miyamoto M, Iwashita S, Yamaguchi S, Ono Y. (2009). Role of nm23 in the regulation of cell shape and migration via Rho family GTPase signals. *Mol Cell Biochem* 329: 175-179.

Munroe SH, Zhu J. (2006). Overlapping transcripts, double-stranded RNA and antisense regulation: a genomic perspective. *Cell Mol Life Sci* 63: 2102-2118.

Myers RB, Srivastava S, Oelschlager DK, Brown D, Grizzle WE. (1996). Expression of nm23-H1 in prostatic intraepithelial neoplasia and adenocarcinoma. *Hum Pathol* 27: 1021-1024.

Nakaya HI, Amaral PP, Louro R, Lopes A, Fachel AA, Moreira YB, El-Jundi TA, da Silva AM, Reis EM, Verjovski-Almeida S. (2007). Genome mapping and expression analyses of human intronic noncoding RNAs reveal tissue-specific patterns and enrichment in genes related to regulation of transcription. *Genome Biol* 8: R43.

Nesterov A, Lu X, Johnson M, Miller GJ, Ivashchenko Y, Kraft AS. (2001). Elevated AKT activity protects the prostate cancer cell line LNCaP from TRAIL-induced apoptosis. *J Biol Chem* 276: 10767-10774.

Nguyen PL, Chen MH, Catalona WJ, Moul JW, Sun L, D'Amico AV. (2009). Predicting prostate cancer mortality among men with

intermediate to high-risk disease and multiple unfavorable risk factors. *Int J Radiat Oncol Biol Phys* 73: 659-664.

Olson MF, Ashworth A, Hall A. (1995). An essential role for Rho, Rac, and Cdc42 GTPases in cell cycle progression through G1. *Science* 269: 1270-1272.

Otsuki Y, Tanaka M, Yoshii S, Kawazoe N, Nakaya K, Sugimura H. (2001). Tumor metastasis suppressor nm23H1 regulates Rac1 GTPase by interaction with Tiam1. *Proc Natl Acad Sci USA* 98: 4385-4390.

Parekh DJ, Ankerst DP, Troyer D, Srivastava S, Thompson IM. (2007). Biomarkers for prostate cancer detection. *J Urol* 178: 2252-2259.

Ponting CP, Oliver PL, Reik W. (2009). Evolution and functions of long noncoding RNAs. *Cell* 136: 629-641.

Reis EM, Nakaya HI, Louro R, Canavez FC, Flatschart AV, Almeida GT, Egidio CM, Paquola AC, Machado AA, Festa F, Yamamoto D, Alvarenga R, da Silva CC, Brito GC, Simon SD, Moreira-Filho CA, Leite KR, Camara-Lopes LH, Campos FS, Gimba E, Vignal GM, El-Dorry H, Sogayar MC, Barcinski MA, da Silva AM, Verjovski-Almeida S. (2004). Antisense intronic non-coding RNA levels correlate to the degree of tumor differentiation in prostate cancer. *Oncogene* 23: 6684-6692.

Reymond A, Volorio S, Merla G, Al-Maghtheh M, Zuffardi O, Bulfone A. (1999). Evidence for interaction between human PRUNE and nm23-H1 NDPKinase. *Oncogene* 18: 7244-7252.

Reynolds MA, Kastury K, Groskopf J, Schalken JA, Rittenhouse H. (2007). Molecular markers for prostate cancer. *Cancer Lett* 249: 5-13.

Salagierski M, Verhaegh GW, Jannink SA, Smit FP, Hessels D, Schalken JA. (2009). Differential expression of PCA3 and its overlapping PRUNE2 transcript in prostate cancer. *Prostate* 70: 70-78.

Saraste M, Sibbald PR, Wittinghofer A. (1990). The P-loop--a common motif in ATP- and GTP-binding proteins. *Trends Biochem Sci* 15: 430-434.

Sardana G, Dowell B, Diamandis EP. (2008). Emerging biomarkers for the diagnosis and prognosis of prostate cancer. *Clin Chem* 54: 1951-1960.

Schalken JA, Hessels D, Verhaegh G. (2003). New targets for therapy in prostate cancer: differential display code 3 (DD3(PCA3)), a highly prostate cancer-specific gene. *Urology* 62: 34-43.

Scheele C, Petrovic N, Faghihi MA, Lassmann T, Fredriksson K, Rooyackers O, Wahlestedt C, Good L, Timmons JA. (2007). The human PINK1 locus is regulated in vivo by a non-coding natural antisense RNA during modulation of mitochondrial function. *BMC Genomics* 8: 74.

Shang X, Zhou YT, Low BC. (2003). Concerted regulation of cell dynamics by BNIP-2 and Cdc42GAP homology/Sec14p-like, proline-rich, and GTPase-activating protein domains of a novel Rho GTPase-activating protein, BPGAP1. *J Biol Chem* 278: 45903-45914.

Sim HG, Telesca D, Culp SH, Ellis WJ, Lange PH, True LD, Lin DW. (2008). Tertiary Gleason pattern 5 in Gleason 7 prostate cancer predicts pathological stage and biochemical recurrence. *J Urol* 179: 1775-1779.

Smith DS, Humphrey PA, Catalona WJ. (1997). The early detection of prostate carcinoma with prostate specific antigen: the Washington University experience. *Cancer* 80: 1852-1856.

Soh UJ, Low BC. (2008). BNIP2 extra long inhibits RhoA and cellular transformation by Lbc RhoGEF via its BCH domain. *J Cell Sci* 121: 1739-1749.

Storz G, Altuvia S, Wassarman KM. (2005). An abundance of RNA regulators. *Annu Rev Biochem* 74: 199-217.

Strausberg RL, Feingold EA, Grouse LH, Derge JG, Klausner RD, Collins FS, Wagner L, Shenmen CM, Schuler GD, Altschul SF, Zeeberg B, Buetow KH, Schaefer CF, Bhat NK, Hopkins RF, Jordan H, Moore T, Max SI, Wang J, Hsieh F, Diatchenko L, Marusina K, Farmer AA, Rubin GM, Hong L, Stapleton M, Soares MB, Bonaldo MF, Casavant TL, Scheetz TE, Brownstein MJ, Usdin TB, Toshiyuki S, Carninci P, Prange C, Raha SS, Loquellano NA, Peters GJ, Abramson RD, Mullahy SJ, Bosak SA, McEwan PJ, McKernan KJ, Malek JA, Gunaratne PH, Richards S, Worley KC, Hale S, Garcia AM, Gay LJ, Hulyk SW, Villalon DK, Muzny DM, Sodergren EJ, Lu X, Gibbs RA, Fahey J, Helton E, Ketteman M, Madan A, Rodrigues S, Sanchez A, Whiting M, Madan A,

Young AC, Shevchenko Y, Bouffard GG, Blakesley RW, Touchman JW, Green ED, Dickson MC, Rodriguez AC, Grimwood J, Schmutz J, Myers RM, Butterfield YS, Krzywinski MI, Skalska U, Smailus DE, Schnerch A, Schein JE, Jones SJ, Marra MA. (2002). Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci USA* 99: 16899-16903.

Sutera VA Jr, Han ES, Rajman LA, Lovett ST. (1999). Mutational analysis of the RecJ exonuclease of Escherichia coli: identification of phosphoesterase motifs. *J Bacteriol* 181: 6098-6102.

Tasheva ES, Roufa DJ. (1995). Regulation of human RPS14 transcription by intronic antisense RNAs and ribosomal protein S14. *Genes Dev* 9: 304-316.

Timmons L, Xu J, Hersperger G, Deng XF, Shearn A. (1995). Point mutations in awdKpn which revert the prune/Killer of prune lethal interaction affect conserved residues that are involved in nucleoside diphosphate kinase substrate binding and catalysis. *J Biol Chem* 270: 23021-23030.

Troyer DA, Mubiru J, Leach RJ, Naylor SL. (2004). Promise and challenge: Markers of prostate cancer detection, diagnosis and prognosis. *Dis Markers* 20: 117-128.

van Gils MP, Cornel EB, Hessels D, Peelen WP, Witjes JA, Mulders PF, Rittenhouse HG, Schalken JA. (2007). Molecular PCA3 diagnostics on prostatic fluid. *Prostate* 67: 881-887.

Yazgan O, Krebs JE. (2007). Noncoding but nonexpendable: transcriptional regulation by large noncoding RNA in eukaryotes. *Biochem Cell Bio* 85: 484-496.

Wu L, Belasco JG. (2008). Let me count the ways: mechanisms of gene regulation by miRNAs and siRNAs. *Mol Cell* 29: 1-7.

Zhou YT, Soh UJ, Shang X, Guy GR, Low BC. (2002). The BNIP-2 and Cdc42GAP homology/Sec14p-like domain of BNIP-Salpha is a novel apoptosis-inducing sequence. *J Biol Chem* 277: 7483-7492.

VITA

Alessandro Kelien Lee was born in Torino, Italy on May 29, 1980, the Son of Emanuela Melchiorre and Joseph Lee. After completing his work at '*Liceo Alessandro Volta*' Scientific High School, Torino, Italy in 1999, he entered The University of Torino, Italy. He received the '*Laurea Magistrale'* [Master of Science] *Summa cum Laude* in Industrial Biotechnology with an experimental thesis in protein engineering from the University of Torino in July, 2005. During his last year in University, he worked as a research intern at NanoBioDesign Ltd, Imperial College London, UK, and prior to joining graduate school he worked as a research assistant at the Institute for Cancer Research and Treatment, Candiolo, Italy, and as a research intern at The University of Texas MD Anderson Cancer Center, Houston, Texas. In August of 2006 he entered The University of Texas Health Science Center at Houston Graduate School of Biomedical Sciences.

Permanent address:

1811 Midnight Ln

Houston, Texas 77047